CONTENTS

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November 1996	Final - September 1995 to March 1996

4. TITLE AND SUBTITLE

Team Performance during Combat Mission Training: A Conceptual Model and Measurement Framework

5. FUNDING NUMBERS

C - F41624-95-C-5011
PE - 62202F
PR - 1123
TA - B2, B3
WU - 06, 01

6. AUTHOR(S)

Alan Spiker
Steven J. Tourville
Denise R. Silverman
Robert T. Nullmeyer

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Anacapa Sciences, Inc.
901 Olive Street
Santa Barbara, CA 93101

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Armstrong Laboratory
Human Resources Directorate
Aircrew Training Research Division
6001 S. Power Road, Building 558
Mesa, AZ 85206-0904

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AL/HR-TR-1996-0092

11. SUPPLEMENTARY NOTES

Armstrong Laboratory Technical Monitor: Dr. Robert T. Nullmeyer (602) 988-6561

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report presents a conceptual model that will guide on-site research on team training effectiveness research at the 58th Special Operations Wing, Kirtland AFB, New Mexico. An objective is to identify the characteristics of effective, mission-ready aircrews to improve training procedures and technologies. Initial focus is on MC-130P aircrew performance during Annual Refresher Training, but the principles underlying effective combat teams should apply to other weapon systems. The report begins by tracing the evolution of team training from its roots in cockpit resource management (CRM) to its subsequent incorporation into combat mission training (CMT) by all three military services. A measurement model is then presented that links coordination processes to team performance and mission outcome. The model assumes that team coordination: is necessary for mission success, is a multidimensional property of the team; and emerges over time in response to programmed training events. Then the report discusses the methodological requirements for conducting CMT-based team performance research, including hypotheses to be tested, the basic experimental design, data collection instruments, key issues, and a multistep strategy for data analysis. We conclude by discussing anticipated research and development impacts, including portable mission readiness assessment tools, improved mission scenarios, and candidate training improvement interventions.

14. SUBJECT TERMS

Cockpit resources management; CMT; CRM; Combat mission training; Flight simulators; Flight simulation; Performance; Performance measurement; Simulators; Team performance; Team training; Teams; Training

15. NUMBER OF PAGES 75

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89) Prescribed by ANSI Std Z-39-18
298-102 COMPUTER GENERATED

TABLES

TABLES

PREFACE

This project was performed for the Aircrew Training Research Division of Armstrong Laboratory, Human Resources Directorate (AL/HRA) by Anacapa Sciences, Inc. and Hughes Training, Inc. (HTI), Training Operations. HTI has a two-year contract (Contract F41624-95-C-5001) with AL/HRA to provide behavioral research support in the areas of combat mission training and mission rehearsal. Anacapa Sciences, Inc. serves as a subcontractor to HTI on this effort, providing scientific and technical support. The contract effort was conducted under Work Unit 1123-B2-06, Aircrew Training Research Support, and the in-house work under 1123-B3-01, Special Operations Forces Aircrew Training and Mission Preparation Research. The AL/HRA Contractor Monitor is Mr. Daniel Mudd; the technical monitor is Dr. Robert T. Nullmeyer.

The authors would like to acknowledge several individuals who contributed generously of their time and support to this report. We would first like to thank LTC Ed Reed, commander of the 58th Training Support Squadron (TRSS), who provided access to the people and state-of-the-art technology associated with the 58 TRSS's Weapon System Trainer/Mission Rehearsal System (WST/MRS). We would also like to acknowledge Mr. John Fuller, HTI's program manager, for his guidance and encouragement throughout this effort. In addition, we would like to thank the HTI Annual Refresher Training instructors at Kirtland AFB and the Air Force Special Operations Command (AFSOC) MC-130P aircrews, whose cooperation and participation in the baseline and on-going efforts are critical to the research.

Finally, we would like to acknowledge the outstanding support of the following experts in Cockpit Resource Management (CRM) who supported site interviews during the project:

Dr. Robert Helmreich, University of Texas at Austin
Mr. John Wilhelm, University of Texas at Austin
Mr. Dave Wilson, Hughes Training, Inc.
Mr. Frank Steiner, Hughes Training, Inc.
Dr. Joseph Weeks, Armstrong Laboratory
Dr. Rick Siems, Armstrong Laboratory

a group, which has a variety of dimensional indicators of its level of development. Third, TEAM coordination is assumed to emerge over time in response to programmed training events and is most reliably manifested under realistic tactical conditions.

Given these overall assumptions, we may delineate the specific objectives that are to be accomplished by the empirical research during the first and second years of the research program, along with some longer-term implications. In *Year 1*, our objective is to identify the characteristics of effective, mission ready aircrews that may be incorporated into training program procedures and technologies. This initial identification is to be accomplished within the context of a correlational design relying on naturalistic observation techniques by highly trained researchers and subject matter experts (SMEs). During *Year 2*, the most promising of these characteristics will be "packaged" into a set of training program variables whose impact will be assessed using an experimental (i.e., control group versus treatment) design. Potential candidate variables might include:

- an enhanced CMT mission scenario
- a program of instructor reinforcement of key CRM behaviors
- use of a geo-specific workstation to deliver situation awareness training
- selective cross-training of key tasks and skills
- just-in-time training of perishable skills
- tailored checklists of required planning activities
- greater emphasis on training team skills versus individual skills

For the longer term, one of the desired strategic outcomes of the research program is the eventual development of a portable, research-based instrument for identifying the team elements contributing to mission ready aircrews—a tool that can be used at the squadron level by commanders, instructor pilots (IPs), and standardization/evaluation (Stan/Eval) personnel to enhance operational readiness and mission effectiveness. In this regard, it is important to understand the implications associated with the assumptions and objectives outlined above. Depending on the behavior, attitude, or cognition being assessed, the research may require the use of observational scoring (counts, timing), researcher-supplied ratings, computer-generated measures, self/instructor ratings, or some index derived from qualitative data. Hence, the measurement methodology that is eventually implemented as part of the research plan must be quite robust and flexible, and include a mix of objective and subjective indices that may be reliably obtained within established resource constraints.

Organization of Report

The remainder of the report is presented in four sections. In the next section, we give a brief background on TEAM training, tracing its evolution within the research community and its relationship to CMT. We then develop a high-level, three-dimensional design to tie CRM and TEAM functional areas into combat mission elements, and finally tie both into behaviors and processes.

Following that, we describe a measurement model of CMT TEAM performance that links TEAM coordination processes to both a TEAM performance and a mission outcome component. In presenting this model, we briefly touch on relevant CRM research by the other services,

focusing principally on a study conducted by the Air Force (Povenmire, Rockway, Bunecke, & Patton, 1989).

Next, we describe the methodological requirements for conducting TEAM performance research within the context of CMT. This description of the required methodology addresses the hypotheses to be tested; the basic design; the data collection instruments to be used; key methodological issues; and a multi-step, detailed strategy for data analysis. While the specific requirements will be illustrated using the MC-130P, the logic of our methodology will apply equally well to all Air Force Special Operations Command (AFSOC) weapon systems. In the final section, we discuss some of the research and development (R&D) impacts that are anticipated from the CMT research program. We will focus on the eventual development of a portable TEAM mission readiness assessment tool (TM-RAT), the specification of improved combat mission scenarios, and the delineation of candidate interventions designed to improve CMT.

EVOLUTION OF TEAM TRAINING

In this section, we offer a brief background on the major concepts and issues that have guided our conceptions of TEAM training within the context of CMT. We begin by discussing how the notion of TEAM training has evolved from its initial pilot-centered, CRM focus to encompass a much wider array of players and issues. As part of this discussion, we delineate the characteristics of a combat TEAM more precisely, focusing on the MC-130P weapon system and highlighting its relationship to CMT.

Team Training Research

From a research standpoint, we recognize that the study of team training has been hampered by the inability to achieve a solid, working definition of the concept (Dyer, 1984). In this regard, development of performance models of team training suffered for many years from the failure to operationally define "teamness" (Salas, Dickinson, Converse, & Tannenbaum, 1992). The lack of an agreed-upon scientific definition of a team has caused confusion in distinguishing teams from groups, where it is assumed that patterns of interaction, methods of promoting cohesiveness, and strategies for reinforcing effective behaviors will be different between the two collectives. Recently, however, there has been a growing consensus that teams are best viewed as lying on a continuum, with a true team defined as a "distinguishable set of two or more people who interact, dynamically, interdependently, and adaptively toward a common and valued goal/objective/mission, who have each been assigned specific roles or functions to perform, and who have a limited life-span of membership" (Salas et al., 1992; p.4). There is no question that however a "team" is defined, SOF aircrews and their associated mission support personnel would most certainly qualify.

But even after identifying a particular collective entity as a team, one must still resolve the issue of how (or whether) team training is to be conceptualized relative to the training required for the skills of individual team members. Training technologists contend there is "considerable overlap" between individual and team training (Andrews, Waag, & Bell, 1992), making the measurement of team performance somewhat problematic. To facilitate the measurement issue, a number of researchers have constructed structural models of team composition and function, with the hope that a priori conceptions will help guide the search for behavioral variables that predict team effectiveness under realistic conditions. Indeed, over the past few years, several models have been put forth in which such factors as task characteristics,

task demands, work characteristics, individual characteristics, and their interactions have been addressed. Unfortunately, an extensive research base has yet to emerge that documents the relative impact of such key variables as stimulus fidelity, response fidelity, performance feedback, and individual skill level on team training design or evaluation (Salas et al., 1992).

Aircrew Performance and CRM

Not surprisingly, the operational community could not wait for the theorists to resolve their disagreements on the defining elements of a team before addressing the vexing problems of aircrew coordination and its likely impact on flight safety. In the 1970s, systematic reviews of airline accident data revealed that numerous errors were committed by crewmembers who were clearly skilled in their individual tasks. The National Transportation Safety Board (NTSB) first noted a need for crew coordination and communication training in 1978 during an investigation of a DC-8 crash in Portland (Clothier, 1991). Surveys and research conducted by NASA suggested that aircrew performance, as measured by accident incidence, could not simply be predicted from technical training, as the more successful crews exhibited consistent patterns of behavior and communication—or coordination—that made them more effective when performing their shared tasks (Lauber, 1987). Analysts surmised that the performance of these coordination-oriented or team tasks made the successful cockpit crews more effective.

From a common sense standpoint, this fact was not surprising, since these shared tasks require common expectations, supportive behaviors, open access to information, and clear communication (Andrews et al., 1992). But programmatically, these skills were not part of any aircrew training curriculum, as the principal focus had always been on acquiring the technical skills necessary to fly the airplane (Prince, Chidester, Cannon-Bowers, & Bowers, 1992).

By 1980, the emerging importance of nontechnical, group-based coordination activities was underscored by several research trends. Even early studies of team performance showed that, while overall team performance is positively related to the individual skills of its team members, the magnitude of the effect is small, suggesting that individual skills are necessary but not sufficient for good team performance (Salas et al., 1992). Thus, one cannot be assured that team performance will be enhanced simply by training team members to master the individual skills required of their duty positions.

Given that the individual-team performance relationship is marginal, what types of shared skills or coordination behaviors emerge that are more strongly associated with outcome? In a landmark study of the effects of workload on aircrew performance, Ruffell Smith (1979) reported that the behaviors that most differentiated crew effectiveness were not technical (i.e., flying related), but lay in the areas of leadership, decision making, and resource management. Thus, it was believed that training these "softer skills," i.e., ones more closely allied with business management and team building, would yield large dividends in terms of increased flight safety, more evenly distributed crew workload, and more efficient communication of information into and out of the cockpit.

Five Generations of CRM Training, Research, and Development

Although these early investigations identified aircrew coordination as an important new area for research and development, it was not immediately clear which specific skills and behaviors should be called out for training. From interviews we conducted with three notable experts in CRM (R.L. Helmreich & D. Wilson, personal communications, September 26, 1995, and A.

Diehl, personal communication, November 1996), it is apparent that the airlines and the US Air Force (USAF) pursued separate, but related, paths of development.

As depicted in Table 1, Helmreich (personal communication) has characterized the shift in CRM emphasis as occurring over five generations. Prior to 1980, the concept of crew coordination or CRM focused on the individual pilot, where course curricula entailed borrowing ideas from social psychology and management to promote more effective leadership and communication. At this early stage, CRM was less a training methodology and more a philosophy in which it was assumed that interventions aimed at promoting group cohesiveness and a more "open communications" cockpit environment would translate into more effective aircrew performance (Helmreich, 1980). While this idea had merit in principle, Helmreich (1980) astutely noted that studies in other social contexts (e.g., groups of aquanauts working underwater) had shown that training designed to improve group cohesiveness per se did not necessarily improve performance. Rather, the relationship between group cohesiveness and group performance was one-way, in which high group cohesiveness is an effect rather than a cause of good group performance (p. 12).

Table 1. Five Generations of CRM Training (from Helmreich, personal communication)

Generation	Types of Data Collected	Primary Thrust	Duty Position Focus
0 (Pre-1980)	Accident Reports	CRM viewed as a loose collection of psychology, leadership, organization, and management concepts	Pilot
1 (Early 1980s)	Aircraft commander attitudes measured against an ideal standard	Attitude awareness; first United Airlines course	Pilot
2 (Mid 1980s)	Line line-oriented flight training (LOFT) checklists (LLC)	Modular approach; Delta Airlines; decision making models, breaking error chain	Flight Deck
3 (Early 1990s)	Later evolutions of LLCs	Team building skills	Aircraft team—pilots, flight attendants, mechanics
4 (Present)	Phase-specific behavioral markers, 4th generation of LLC, operations data collected to guide scenario development	Data-driven cultural barrier assessments; address airline-specific problems; in use by Southwest Airlines	Entire team—pilots, flight attendants, mechanics, air traffic controllers, schedulers, etc.

With the launching of the first CRM course in the early 1980s, United Airlines attempted to improve the attitudes of the individual pilot in order to promote more communication and information-sharing in the cockpit. The emphasis here was on developing tests that would identify certain personality types most likely to resist coordination concepts, and then develop training programs to foster a more communicative environment among those individuals. As such, CRM training impacts were seen as operating mainly on select pilots, to "fix" the ones most likely to resist information from co-pilots in time-critical, high workload situations (Helmreich, personal communication).

By the mid-1980s, the airlines' CRM thrust had expanded beyond the individual pilot to include everyone on the flight deck (flight engineers, co-pilots). It was assumed that the need for

improvement in coordination skills existed in all flight deck personnel. As part of this approach, Delta Airlines developed a CRM course designed to address weaknesses in human information processing and decision-making, in which techniques for information acquisition and dissemination were taught to enhance crew performance. In addition, checklists were developed so that on-board evaluators could rate pilots' proficiency in performing various CRM skills and behaviors (Clothier, 1991).

By the early 1990s, CRM was being promoted as a set of specific skills and behaviors in which building effective teams was the desired goal. CRM courses were divided into sections, with each section devoted to training and reinforcing skills in a particular cluster. These included (a) communication processes and decisions (e.g., listening, conflict resolution), (b) team building and maintenance (e.g., leadership, followership), and (c) workload management and situational awareness (Gregorich & Wilhelm, 1993). Importantly, the focus of CRM was expanded beyond the cockpit, to include all members of the aircraft team (i.e., flight attendants, mechanics). It was believed that by having everyone in the "information loop," there would be a greater chance of identifying any and all problem factors, thereby breaking the error chain that characterizes most major accidents (Diehl, 1992).

For some airlines, CRM training has an expanded focus beyond the aircraft itself to include all members of the airspace management team such as air traffic controllers, schedulers, and the like. There has also developed a more problem-oriented approach, in which each airline explores the particular CRM-related problems that plague their operations rather than taking a more global, industry-wide perspective. In this regard, there is a greater emphasis upon collecting substantial amounts of quantitative data derived from ratings of aircrew coordination effectiveness provided by check airmen who fly with the crew. Along with this problem-oriented approach, the exporting of CRM concepts to airlines outside this country has prompted the development of checklists which identify "behavioral markers" (e.g., acknowledging receipt of a communication) that give an indirect indication of the fundamental, underlying CRM subprocesses. A number of airlines, both domestic and foreign, are collecting rating-based CRM skill marker data across different phases of flight and operating conditions in order to leverage large-scale data against specific problems of interest (Law & Wilhelm, 1995).

In reviewing the status of CRM within the airline industry, Gregorich and Wilhelm (1993) noted that efforts at promoting CRM acceptance and improved targeted attitudes have been successful and the learning of key CRM concepts has been demonstrated. However, these improvements in CRM attitudes and awareness have yet to be linked directly to behaviors in flight simulators or line operations. It is rather disappointing that despite more than ten years of CRM development and promotion, a direct association between CRM training and aircrew performance in the airlines has yet to be established.

Air Force CRM and Combat Mission Training

Within the USAF, applications of CRM concepts to training programs in general, and CMT in particular, have exhibited a similar evolutionary trend. This progression is depicted in Figure 1, in which CRM training was originally focused on the individual aircraft commander. Over time, coordination concepts have been expanded to include other crewmembers, the larger mission crew (e.g., maintenance), and ultimately, the entire combat mission team (Nullmeyer, 1995). Initially, USAF CRM course offerings mirrored those of the airlines, in which pre-existing attitudes were assessed, management-oriented techniques presented, and case studies

from weapon system-specific accidents described (D. Wilson, personal communication, September 28, 1995).

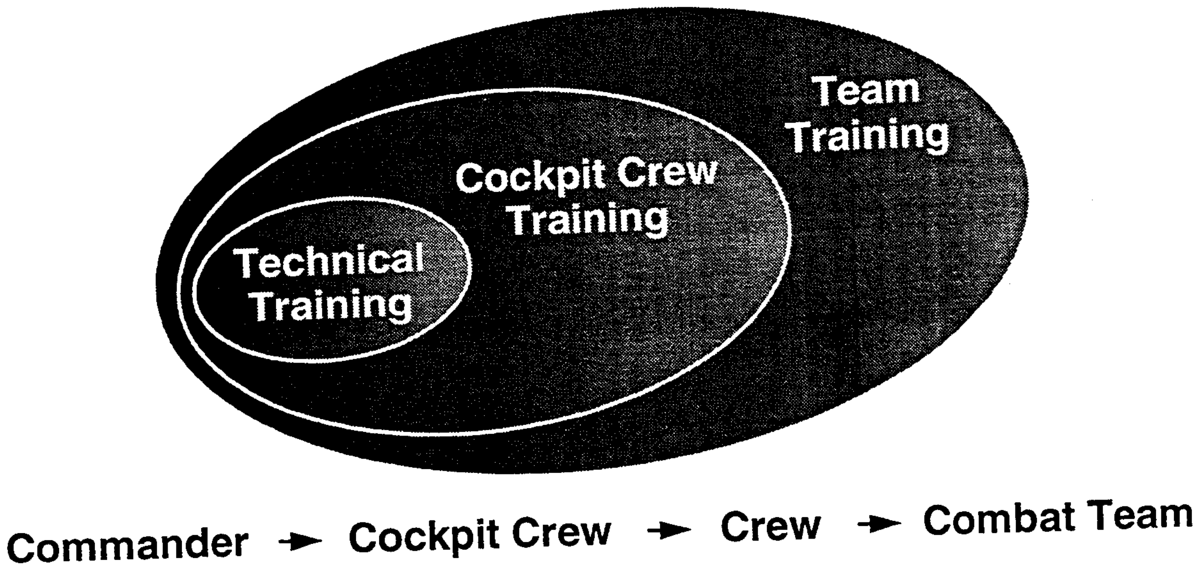

Figure 1.
Evolution of Team Training in the Air Force (Nullmeyer, 1995).

Over time, a modular approach (like Generation 2 for the airlines) was incorporated into many USAF CRM programs whereby five types of management skills are covered—attention management, crew management, stress management, attitude management, and risk management (Diehl, 1992). Thus, like the commercial airline industry, the USAF focus on CRM training has moved away from the "touchy feely" preoccupation with attitudes and toward the development and reinforcement of more tangible skills and behaviors. This movement represents a realization that pilots tend to resist the teaching of "soft," non-technical skills as well as a belief that more effective aircrew performance will result from more tangible behaviors (D. Wilson, personal communication, September 28, 1995).

Besides modularization of CRM concepts, the USAF approach to training has entailed an identification of more specific functional areas that warrant training and reinforcement. This trend is consistent with the airline's evolution toward functional specification during the transition between Generations 2 and 3. A good example of this view is represented in the CRM pyramid of critical success factors, depicted in Figure 2. While serving a number of purposes, the pyramid is often used as a pedagogical tool to illustrate to aircraft commanders the quantitative involvement of different factors in Class A accidents. Analyses of USAF accident data have attributed approximately one in five accidents to either violations of cockpit procedures and rules or to loss of situation awareness (SA). Other factors shown inside the triangle include command authority, communications, workload management, resource utilization, decision-making, and operating strategy (i.e., flying the aircraft). Importantly, whether one truly believes the incident rates attributed to the individual critical success factors, the mere act of identification serves to improve the specificity of CRM training and, presumably, its overall utility to aircrews (D. Wilson, personal communication, September 28, 1995).

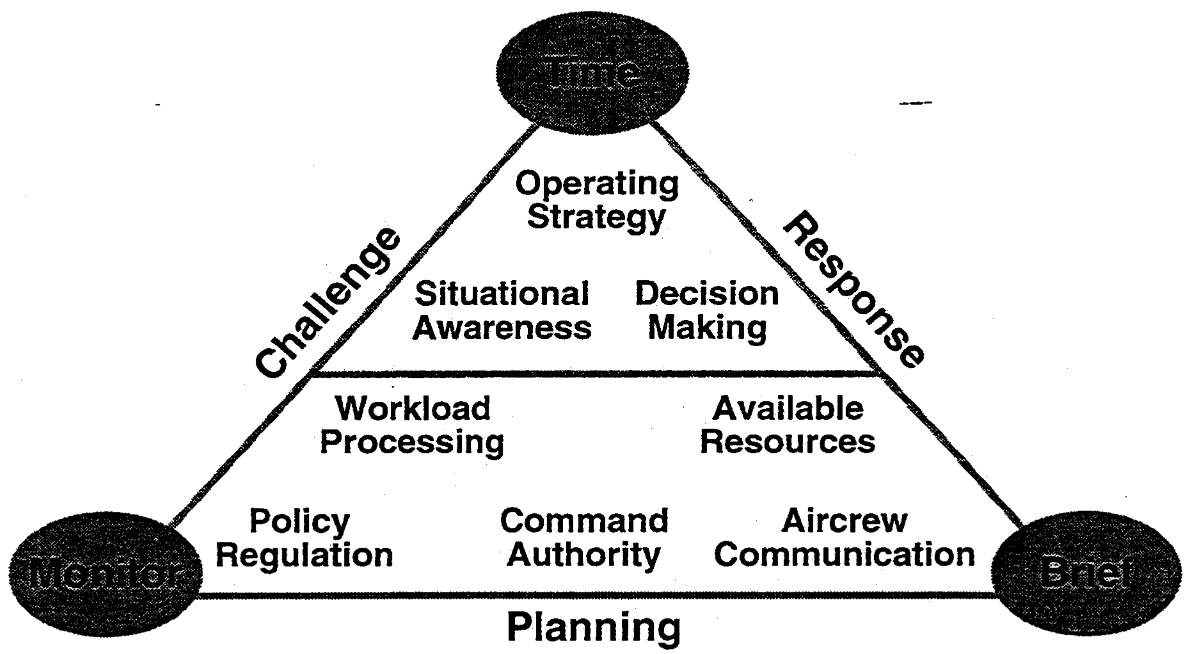

Figure 2.
USAF CRM Pyramid and Critical Success Factors.

By our definition, a TEAM is composed of those players whose information, actions, and decisions impact the aircrew's mission in some way (Nullmeyer, 1995). Besides the aircrew itself, this encompasses such mission participants as intelligence, weapons and tactics (W&T), logistics, weather, airborne command and control, air traffic control (ATC), maintenance, as well as the ground "customers" (Rangers, Seals, etc.) being supported by the AFSOC aircrew. During CMT, the actions of this extended family may be role-played by the instructors, whose inputs are scripted to promote standardization of training. Such scripting is ideal from a research standpoint, as it reduces unwanted stimulus variation across subject-crews.

To illustrate the wide diversity of roles and responsibilities present in a typical SOF mission, Figure 3 depicts some of the major players in an MC-130P mission whose functions should be represented and/or simulated. As shown in the center of the figure, the standard MC-130P aircrew consists of the aircraft commander (AC), co-pilot (CP), flight engineer (FE), left and right navigator (LN, RN), and communications system operator (CSO)[1] Surrounding this box we have depicted the key players who will provide necessary supporting information to the aircrew during mission preparation and/or execution. This includes the functions of intelligence (Intel), W&T, ATC, maintenance, logistics, and weather (Wx). In addition, there are other members of the combat mission team whose tactical actions directly affect the aircrew, such as any Special Forces (SF) on-board "customers" as well as other AFSOC aircraft that will be participating in the mission (e.g., a helicopter that is to be refueled by the MC-130P). Besides the supporting and tactical players, the TEAM also includes those individuals and functions that provide a command and control role, such as the squadron/wing command leadership (CC/DO/ADO) and the airborne command, control and communication (ABCCC) aircraft.

[1] We have omitted from this discussion a seventh crewmember, the loadmaster (LM). There is at present no simulation capability at the 58 TRSS to support the training of LM technical and CRM skills. While the LM receives the CRM classroom training on Day 1 with the rest of the crew, he is trained on his tasks at a separate site for the rest of the week and does not participate in the Day 5 combat mission scenario.

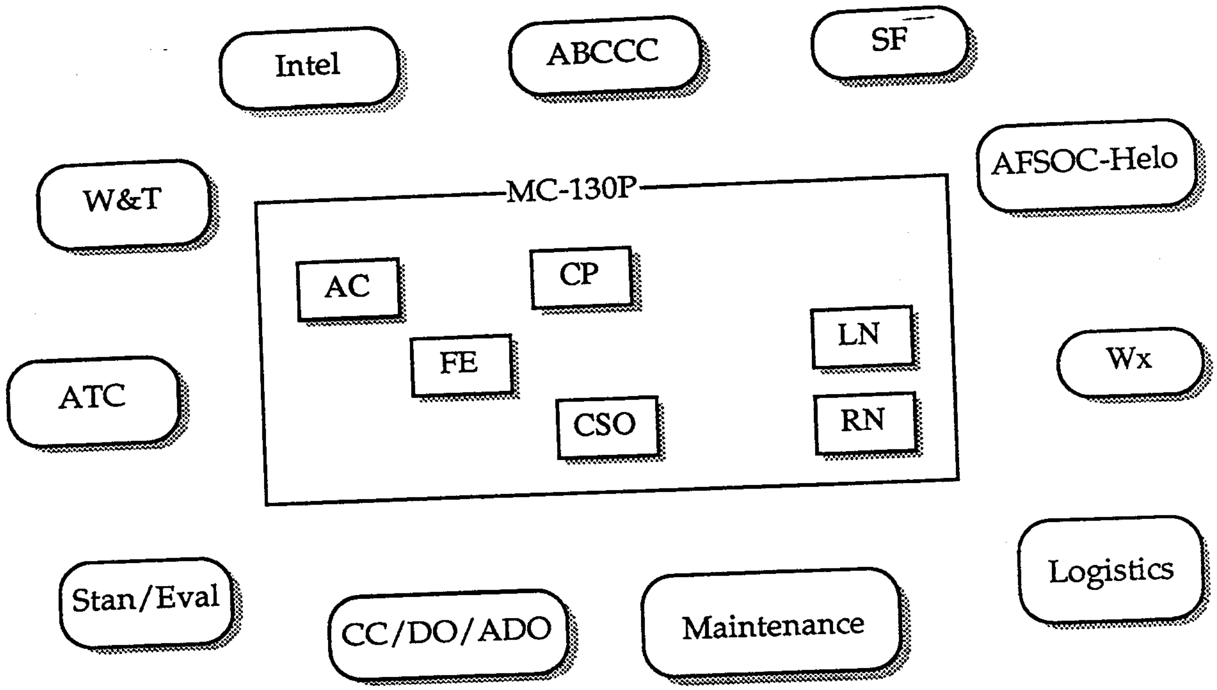

Figure 3. Schematic of the MC-130P Combat Mission Team.

As we undertake the study of TEAM performance during CMT, it is important to have a clear idea of the role that CMT plays within USAF training in general, as well as the subrole that CRM training plays within CMT. Several recent AL/HRA reports have described the standard three-phase progression of training that all aircrews undergo (Bruce, Killion, Rockway, & Povenmire, 1991; Bruce, 1989), consisting of initial qualification (Phase I), mission qualification (Phase II), and continuation training (Phase III). Strictly speaking, CMT can occur during any of these phases, as it most properly refers to the training of tactical tasks, behaviors, and processes that are required for success in combat. Thus, CMT does not enjoy the status of a specific phase of training, but rather, is considered to underlie much of what should go into a tactically-oriented curriculum. However, it is fair to say that the majority of what would be considered CMT occurs during continuation training, of which the aircrew's one-week of annual (simulator) refresher training is a significant part.

At a conceptual level, it is instructive to consider CMT as an integration of multiple training events that combine to transform an aircrew into a combat mission ready TEAM. For Annual Refresher Training, this notion is illustrated in Figure 4, in which CMT is depicted as the joint intersection of CRM academics and simulator evaluation, technical training (e.g., emergency procedures, systems), and combat tactics (D. Wilson, personal communication, September 28, 1995). The latter includes a wide range of events, such as threat recognition, deployment of expendables, mission planning, aerial refueling, low-level navigation, and clandestine insertion and extraction. But importantly, it is this added emphasis on training and maintaining combat tactical skills that makes the study of TEAM performance in the military setting so different from its commercial counterpart. As we shall see, the desire to capture aircrew coordination processes underlying combat tactics requires the development of customized measurement procedures, ones that cannot simply be borrowed from the airline industry.

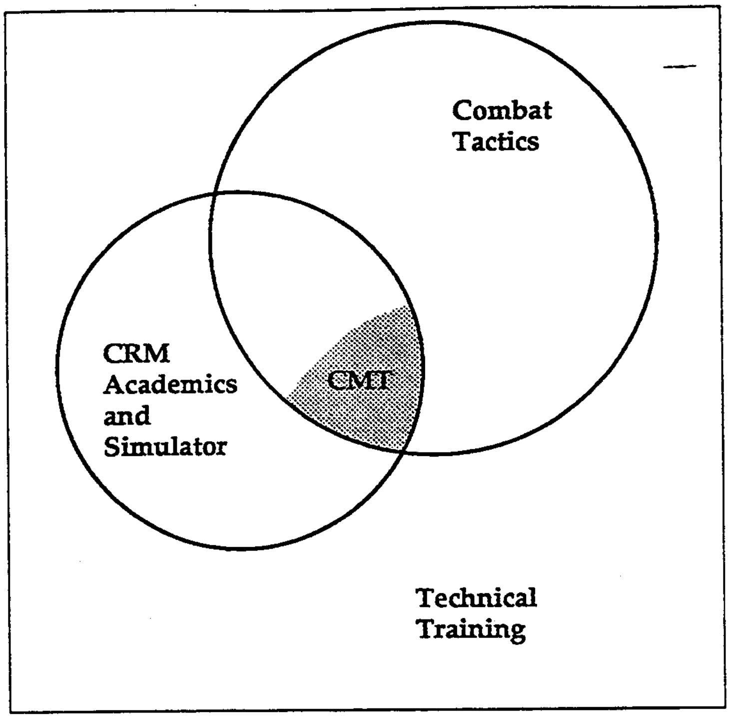

Figure 4.
Combat Mission Training and Annual Refresher Training.

Regulatory Environment

Given our emphasis on CRM and TEAM performance measurement and our interest in the MC-130P weapon system, two regulatory requirements are particularly important: Air Force Instruction (AFI) for CRM Training (AFI 36-2243) and the mission definition portion of AFSOC Regulation (AFR) 51-130. The first document describes the methods that can be used to train and measure CRM task performance. The second document outlines the tasks to be performed.

The AFI is a particularly important document for our purposes, as it establishes the requirement that each major command (MAJCOM) measure CRM. The three stated objectives of the AFI are to: (a) maximize operational effectiveness and combat capability, (b) prevent accidents and incidents, and (c) improve efficiency of all forms of training. The "MAJCOMs and FOAs (field operating agencies) will periodically review on-going human factors research for potential training applications" (p. 6). With regard to CRM content, the AFI dictates that the core curriculum cover eight concepts: SA; group dynamics; effective communications; risk management and decision-making; workload management; stress awareness and management; mission planning, review, and critique strategies; and human performance. The latter area includes SA elements, technological aspects of SA (controls, displays), enhancing SA, loss of SA, and error source identification and error prevention.

MAJCOMs must comply with these requirements during each phase of training, including Annual Refresher Training. In addition, the document broadens the measurement domain of CRM-related performance: "Although CRM programs are mandatory for aircrew members and have historically been geared toward the operational flying environment, the *potential exists to adapt fundamental program principles to any task or functional area requiring*

cooperative or interactive time critical efforts (p. 1)." This latter stipulation opens the door to look at TEAM behaviors that go beyond traditional cockpit boundaries.

The AFI establishes some definite, albeit broad, boundaries around the content areas within which our own data collection efforts will be couched. From an evaluation standpoint, the AFI states that training units must perform assessments using the following program measurement metrics: (a) establish tracking of inflight stan/eval trends data with CRM implications; (b) track and assess Class A, B, and C mishaps; and (c) analyze and track CRM-related implications based on student critiques.

In addition, the AFI stipulates that two methods for CRM program evaluation are acceptable: (a) over-the-shoulder observation by evaluators, qualified line observers, and MOST instructors to collect skill performance data, and (b) the conduct of "derivative" (before/after surveys of trainees) crew surveys. This stipulation is quite significant from a research standpoint as it expands the range of acceptable measurement options over and above the three primary metrics. This is particularly important for measuring performance in a simulator, for which accident data and check-ride ratings are not applicable.

In reviewing the detailed information in the AFI, it is important to note the following caveat: "trainers may use personality and attitude measures to provide self-awareness feedback to participants, tailoring them to assess strengths and weaknesses. Such tools will be used for classroom purposes only and will not be part of any program effectiveness measurement system" (p. 8). Thus, personality-oriented tests should only be used for internal assessments and not for any external evaluation. As a result, we have elected to exclude the use of traditional personality tests as a source of CRM data for our model of TEAM performance measurement.

In sum, the AFI is truly an essential element of CRM training, as it highlights the importance of SA and mission planning in CRM, expands the scope of CRM concepts to the tactical environment, permits over-the-shoulder observation as a legitimate data collection technique, and requires the conduct of a pre-post crew CRM survey. These features will be discussed in detail in the next section, when the conceptual model of measurement is presented.

The second document, AFSOC Reg 51-130 (Flying Training), serves as the primary source to determine which mission events are to be trained and whether they are being accomplished to a satisfactory extent. By identifying the critical mission training events, this document forms the basis for, and establishes the validity of, the mission scenario that is used to conduct CMT in the weapon system trainer (WST). For the MC-130P weapon system, Table 2 summarizes the key events, their requirements, and associated criterion indices. Assessments of scenario adequacy will be based on direct comparisons of presented/trained events to those stated in Table 2.

Table 2. Key MC-130P Training Events, Requirements, and Criterion Indices

Key Training Events	Requirements	Criterion Indices
1. Night Vision Goggle (NVG)-Low-level	realistic threat scenario fly route in conjunction with an aerial refueling control time/time on target/time of arrival (ARCT/TOT/TOA) to an actual or simulated AR, airdrop, or airland event includes at least one aircraft A/C defensive maneuver	minimum time enroute is 1 hour
2. Aerial Refueling (AR)	navs must use electronic aid to direct A/C FEs must complete AR checklist LMs must have actual contact by a helo	timing criteria is on-time to 1 min. late for ARCT
3. Airland Operations	(1) Maximum Effort Landings (2) Maximum Effort Takeoffs (3) NVG Landing, Tax, Takeoff (4) Self Contained Approach (SCA) (5) NVG Go-Around (6) Infil/Exfil	TOT is +/- 30 sec. landing zones (LZs) must be 3500 ft or less pt of touchdown must be within the applicable zone covert lighting A/C stops at prebriefed location immediate on/offloading of personnel/equipment complete infil/exfil checklist while on/offloading personnel
4. Airdrop Operations		TOT criteria is +/-30 sec. drop within 300m of aim point
5. Communications Events	secure voice authentication encode/decode HAVE QUICK	
6. Formation	30 min. of wing time	
7. Other Events	Wx Penetration Inflight Refueling Hot Refueling	
8. Electronic Warfare Operations	ground radar event engagement with ground or shipboard SAM/AAA radar site or simulator each event will include a min. of 15 min. activity expendable events (chaff, flares)	

Functional Analysis of CRM

In the following discussion, we first review some of the functional areas that have been singled out for study by other researchers in the CRM field. We then define and describe the five functions that we have elected to pursue, based on their relevance to CMT and the AFSOC mission environment.

CRM-Related Functions

At the outset, it should be noted that there are two basic strategies that one may take in identifying CRM or crew coordination functions, with different research groups tending toward one end or the other. On the one hand, it is possible to identify rather large-scale, global variables that have broad appeal but which lack specificity for immediate implementation in a particular CMT environment. Alternatively, one may focus on identifying a particular context-specific behavior (or functional class of behaviors) that has limited generalization but which is more amenable to immediate training and/or reinforcement. Clearly, each approach has its strengths and weaknesses.

As an example of the former, Stout, Prince, Baker, Bergondy, and Salas (1992) outlined seven CRM meta-skills that should be applicable to all airframes and mission environments: decision making, assertiveness, mission analysis, communication, leadership, adaptability/ flexibility, and SA. In an attempt to assess the generality of these concepts, the researchers asked representative operators from a host of fixed-wing (e.g., F-14, T-44) and rotary wing (e.g., MH-53) airframes to rate the overall importance of each meta-skill to mission success. Although analyses of the ratings revealed some agreement (e.g., SA is rated high for all airframes), it is not clear from these data whether other dimensions would have been rated as important, and whether more precise delineation at the performance standard level would have been even more well-received by the operators.

Following the second strategy, some researchers have attempted to identify more specific CRM behaviors (e.g., prompting a crewmember for a particular piece of information) whose occurrence can be "induced" in a simulator (or aircraft) and then reinforced by an instructor (Oser, McCallum, Salas, & Morgan, 1989). With this approach, the typical method involves a systematic review of critical incidents reported by large numbers of operators, followed by a checklist of the most frequently reported events. In one application of this approach, some 50 critical behaviors were identified as being related to aircrew coordination, where these behaviors were grouped according to the seven meta-skills listed above (Fowlkes, Lane, Salas, Franz, & Oser, 1994). Examples include repeating information, acknowledging communications, cross checking information sources, and identifying alternatives. An advantage of this approach is the reduced reliance on SME judgment for reliably identifying when these behaviors do occur. On the other hand, it remains to be seen whether this approach can be used to structure a CMT program in which prescriptive scenarios determine when these behaviors should occur.

The Army has attempted to combine the two approaches for CRM training for rotary wing operations. Specifically, they identified 13 "Basic Qualities" (BQs) that, collectively, constitute aircrew coordination. These qualities are viewed just like any technical task that must be trained to some criterion level of skill, and which must be evaluated by an IP before the student may advance to the next level. Rating forms are developed for use by the IP and entered into the student's training form. The 13 skills are crew climate, planning/rehearsal, decision techniques, workload management, unexpected events, information transfer, SA, communication acknowledgment, information sought, cross monitoring, information offered, advocacy/assertion, and after action reviews (Zeller, 1992). This breakout has proven useful for IP evaluations, with student pilots exhibiting the proper improvement in BQs over training flights after having mastered more rudimentary technical tasks. However, researchers have found the functional breakouts too general to permit anything other than global judgments (Grubb, Simon, Leedom, & Zeller, 1993; Leedom & Simon, 1993).

Five Functional Areas

As is evident from the preceding discussion, the strategy one uses to select CRM functional areas will have a major impact on the research directions that will be followed. In the approach outlined below, we will emulate the Army's hybrid approach of defining functional areas based on general considerations of specific aircrew tasks. This is in contrast to defining functional areas based on either global dimensions of performance (e.g., CRM meta-skills) or situation-specific considerations (e.g., critical behaviors). In particular, since our ultimate goal is to elucidate the coordination processes that contribute to effective aircrews and good TEAM performance, we selected functional areas based on their: judged relevance to the AFSOC mission environment and previously reported operational problems, appropriateness to the high levels of experience and motivation of many MC-130P aircrews, applicability to CMT and Annual Refresher Training, and amenability to measurement by outside observers. In addition, and where possible, we attempted to derive functional areas that make contact with some of the more global dimensions that have been identified by other researchers.

Based on these considerations, five CRM functional areas were identified to receive further research emphasis: Function Allocation (FA), Tactics Employment (TE), Situation Awareness (SA), Command-Control-Communications (C3), and Time Management (TM). An operational definition for each functional area is provided in Table 3. In reviewing the scope and content of these areas, it is evident that some of the traditional dimensions of CRM—such as leadership, group cohesiveness, personalities, etc.—have been omitted. We readily acknowledge that the five areas that we have identified by no means encompass the entire content domain of what would properly be considered TEAM coordination, and that other factors are worthy of study in their own right. Rather, our focus is on identifying CRM functional areas whose performance have the most direct links to training-related processes and training interventions.

Table 3. Definitions of Five CRM Functional Areas

Functional Area	Definition
Time Management (TM)	Involves the ability of the combat mission team to employ and manage limited time resources so that all tasks receive sufficient time to be performed correctly, and critical tasks are not omitted
Tactics Employment (TE)	Includes all analytic activities necessary to avoid or minimize threat detection or exposure, and to successfully coordinate complex mission events and multiple mission objectives
Function Allocation (FA)	Includes the division of crew responsibilities so that workload is distributed among the crew, avoiding redundant tasking, task overload, and crewmember disinterest or noninvolvement. Tasks should be allocated in such a manner that crewmembers are able to share information and coordinate responsibilities.
Situation Awareness (SA)	Entails maintaining an accurate mental picture of mission events and objectives as they unfold over time and space. Emphasis and analysis are placed on three levels of SA (perception, integration, and generation; Endsley, 1995) and their impact on team coordination.
Command-Control-Communications (C3)	Encompasses those activities required to involve external parties in the mission and to maintain communications with these external team members; communication within the crew; and controlling the sequence of mission events according to the mission execution plan.

In a subsequent section, we will discuss the methods and procedures that will be used to operationalize each functional area. However, before we turn to that topic, we must first pinpoint in time where these functions occur during a mission.

Mission Scenario Elements

At a process level, our approach will be more focused on the mission scenario being flown and less on the particular events that occur. As such, it is not as scripted or controlled as some approaches. Yet, we believe that a scenario orientation is more appropriate when attempting to conduct empirical research within a fluid training environment in which instructor turnover is high, training schedules change from week to week, and a "frozen" set of events is hard to ensure. This more general approach to process also permits expansion to distributed interactive simulation (DIS) environments where other agents control the stimulus-events.

In particular, our process approach identifies a smaller set of discrete phases from the mission scenario. Within each phase, there are multiple opportunities for one or more crewmembers to engage in the behavioral processes that fall within the five functional areas described in the previous subsection (i.e., SA, FA, TM, TE, C3). For the MC-130P weapon system, process measures are provided across five CMT phases. These phases are defined in Table 4.

Table 4. Five Mission Phases for an MC-130P Mission Scenario

Mission Phases	Description
Mission Preparation	The objective is to conduct pre-mission planning and briefing activities that allow sufficient preparation of a comprehensive mission execution plan. This plan will be prepared with considerations for a medium threat environment, all major mission events and activities, and mission operations procedural constraints.
Low-level and Tactics	The objective is to conduct NVG low-level flight enroute to specific mission events using proper tactical mission management procedures (altitude, airspeed, terrain masking, routing) for a medium-threat environment.
Aerial Refueling	The objective is to successfully conduct tactical in-flight aerial refueling of multiple MH-53J Pave Low helicopters within prescribed time, course, and altitude constraints in a medium threat environment.
Airdrop	The objective is to successfully conduct computer aided release point (CARP) airland of special forces personnel within prescribed time, course, and altitude constraints in a medium-threat environment.
Infiltration/Exfiltration	The objective is to successfully conduct covert infiltration and/or exfiltration at multiple tactical landing sites for transload purposes within prescribed time, course, and altitude constraints in a medium-threat environment.

Within each phase, there are a multitude of initiating tactical events that the aircrew and mission team must handle. Accordingly, there will be multiple occurrences of many similar events, permitting an opportunity to aggregate data across repeated measures to construct composites indicative of TEAM coordination processes.

The success of our approach hinges most critically on the quality of the mission scenario that is flown. In this regard, it is important that the scenario induce the particular types of coordination behaviors and processes that we want to observe. The phases described in Table 4 are based on a complex, high workload scenario that was developed by the second author (a former navigator and instructor) for the final WST-CRM mission to be flown on Day 5 of Annual Refresher Training.

Specific guidelines have been drafted by NASA and the FAA for a CRM-oriented mission training (Prince, Oser, Salas, & Woodruff, 1993). While it is recognized that scenario design is not an exact science, the guidelines nevertheless offer some rules of thumb to which prudent researchers and trainers should adhere. Guidelines most applicable to the present research program include:

1. A scenario should contain four major phases—briefing, preflight planning and activities, the flight segment, and debriefing. This guideline implies that only some of the scenario occurs in the simulator itself.
2. The scenario facilitator should give crewmembers information about the scenario they are about to undergo.
3. The scenario should build in the opportunity for crewmembers to demonstrate the behavior of interest more than one time, thus increasing the opportunity for reinforcement and feedback.
4. Realism should be built into the scenario, such as having a real-world database and operationally relevant threats.
5. Include all crewmembers in the scenario, either by actual participation or by role playing. This notion can be logically extended to include the entire team.
6. Add irrelevant communications and possible unforeseen events that might occur in the actual mission.
7. Use a modular format in building scenarios, thus allowing components of the scenario to be enhanced or exchanged without having to rebuild the entire scenario from scratch.

As we begin to consider the methodological requirements for capturing the coordination processes underlying TEAM performance in more detail, we will explore alternative methods of extracting higher order process measures from lower order observations, through an approach we call "derived indices." This will be discussed in more detail in the section where our data collection instruments and measurement procedures are discussed.

CRM Behaviors/Processes

Alternative Approaches

An information-oriented view of process has, not surprisingly, led some researchers to focus on one key functional area—communication. In this vein, a typical approach to process modeling has been the systematic decomposition of mission segments into event-action-outcome sequences. This has been attempted in several forms. For example, Predmore (1991) analyzed the verbal interactions obtained from the cockpit voice recorder during the minutes after United Flight 811 lost a cargo door enroute to Honolulu as well as from the accident in which United Flight 232 lost engine and hydraulics over Sioux City. A verbal coding scheme was developed that assigned communications to classes of "thought units." At a process level, thought units are utterances that deal with a single thought, intent, or action. Independent raters of the transcripts created a coding scheme by classifying thought units as either command-advocacy, inquiry, incomplete-interrupted, or reply-acknowledgment. Each unit was recorded in terms of its speaker, target, and time onset.

Although an analysis of the qualitative pattern of different units was interesting, it was the quantitative trends that proved most dramatic. In particular, Predmore (1991) found that the rate of thought units expressed in the cockpit increased dramatically following the occurrence of a stressful event. For example, in Flight 811, the crew expressed an average of 5 thought units per minute prior to loss of the cargo door and almost 19 units per minute following door loss. Further analysis of the trends revealed that in unsuccessful responses to emergencies, cockpit communication patterns develop in which one crewmember dominates, insufficient communications are transmitted/received from outside the cockpit, and a large percentage of communications are either interrupted or incomplete. While such process data fall short of achieving a performance standard for crew coordination, they certainly are indicative of patterns of communication that are associated with more or less successful outcomes.

Another interesting attempt at process modeling is found in an unpublished study by Schmidt (1987). In this study, Schmidt developed 25 separate categories of communications behavior using an interaction process analysis methodology. Examples include such categories as question, opinion, acknowledgment, agreement, disagreement, and command. This scheme was applied to the coordination behaviors of C-130 aircrews who received mission oriented simulator training, in which observers viewed videotapes of the crews and recorded the timing and frequency of each type of communication. These distributions were then correlated with independent assessments of "successful" and "unsuccessful" crews. Importantly, Schmidt found several communication patterns that were associated with the more successful crews. These included having a greater frequency of communication, fewer communications "left open," and use of problem solving as a primary source of conflict resolution.

The AFSOC model of CRM aircrew coordination provides a third alternative time/event-based approach, wherein the process is decomposed into the planning —> challenge—> response cycle depicted outside the pyramid in Figure 2. As noted previously, this approach is effective from an instructional standpoint, as it allows trainers to place previous operational problems into distinct categories to receive more focused intervention (D. Wilson, personal communication, September 28, 1995). On the other hand, this delineation is not immediately amenable to experimental testing since the stimulus conditions that must be present to induce a particular behavior, positive or negative, are not specified.

While these event-centered attempts at modeling the process aspects of crew coordination have the potential to provide truly useful insights concerning TEAM effectiveness, they have two drawbacks that precluded their further consideration for the present work. First, they tend to require extensive use of videotaping or audio-taping. Such recording is highly resource-intensive, and while it can serve as a rich source of data, it is not always practical when working on a not-to-interfere basis with aircrews who are receiving CMT. Second, the training delivery environment is quite fluid, such that reliance on having events remain consistent over successive training sessions is unlikely, reducing the expected reliability of the associated measures. For these reasons, we have elected to pursue an alternative approach to capturing TEAM process data, described below.

Present Approach

Our proposed approach to assessing TEAM coordination processes is shown in Figure 5. This figure shows a cube that depicts the relationships among the three components of our research design: CRM functional areas, mission scenario elements, and crew/team behaviors and processes. Each face of the cube contains a five-by-five matrix in which the five CRM functional areas (SA, C3, TM, TE, FA) are factorially related to the five mission scenario elements (mission

preparation, low-level and tactics, aerial refueling, airdrop, infil/exfil). The 25 cells of this matrix comprise the "content shell" within which we will be attempting to observe, measure, and record the key crew and team behaviors/processes that are most directly associated with effective coordination.

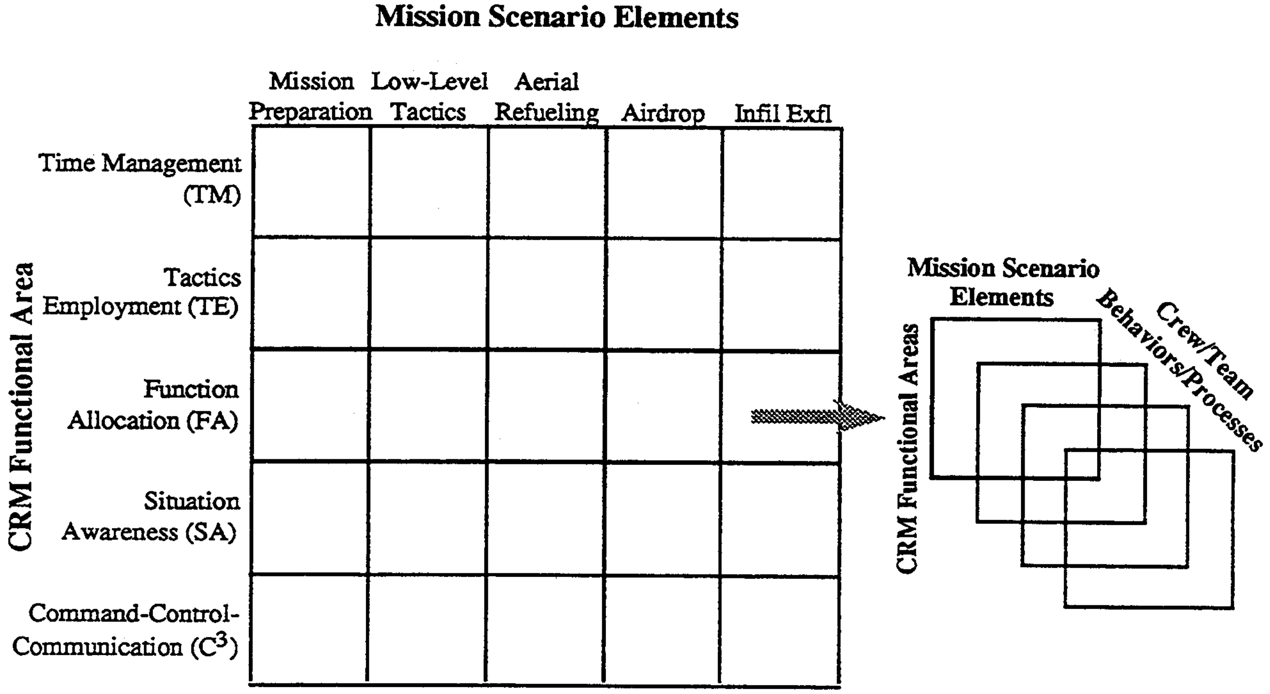

Figure 5.
Relationship of CRM Functional Areas to MC-130P Operations.

Within each cell, the research team will need to construct specialized measurement instruments and recording sheets that will cue observers regarding the specific behaviors—in some cases by individual crew position and in other cases by the aircrew as a whole—that reflect either strong or weak TEAM coordination. The behaviors or behavior classes will be identified based on subject matter knowledge, review of critical incidents, and interviews with current instructors, commanders, and squadron personnel. By locating each behavior within a particular cell, the research team will have both the topic domain and time frame identified, making the procedural aspects of measurement more manageable. Before describing the contents of our data collection instruments more specifically, we first discuss some research that has been conducted in the area of TEAM mission performance within the context of a conceptual measurement model.

MEASUREMENT OF COMBAT MISSION TRAINING TEAM PERFORMANCE

In this section, we discuss some of the procedures that researchers can use to determine empirical relationships between CRM-related processes and indices of TEAM effectiveness. We begin by presenting a conceptual model that we believe useful for understanding the interactions among training events, CRM processes, TEAM performance, and overall mission outcome. Using the model as an organizing framework, we next briefly summarize some of the CRM performance studies that have been conducted by the Army, Navy, and commercial airlines. We

conclude by describing the innovative methodology used in the Povenmire, Rockway, Bunecke, and Patton (1989) USAF study of aircrew coordination. An appreciation of the significance of the Povenmire et al. study is particularly important here, as it lays the foundations for the methodological requirements discussed in the next section.

Conceptual Model

Our conceptual model of TEAM-CMT performance measurement is presented in Figure 6. As will be shown, the model is intended to capture the basic logic of our measurement approach and it outlines the salient methodological features that will characterize our research. In this vein, the model serves as a blueprint that suggests where researchers should develop instruments to collect CRM-related data during CMT. The model is also useful as a means of seeing the "big picture" within the CRM domain, to help interpret the findings from related studies in the field in a more organized fashion.

At the outset, we readily acknowledge that the conceptual model is an oversimplification of the CRM domain as it omits many of the links that undoubtedly exist among variables operating at multiple levels. We have done this in the interest of clarity of presentation so we may focus on the variables of most immediate interest. In addition, the concepts and the arrows linking them should be read from left to right, to be consistent with an implicit timeline of training activities occurring throughout the week of Annual Refresher Training.

Starting from the far left, the figure depicts three modules with arrows feeding into TEAM coordination processes, the primary focus of our research. The first two, Crew Background and Baseline Attitudes, reflect the fact that, going into training, aircrews will vary in terms of their background experiences (e.g., squadron affiliation, number of similar CMT missions conducted, hours flown together as a crew) and attitudes toward CRM principles and CRM training. With regard to the latter, analysts have posited that crew attitudes towards CRM academics and simulation sessions (Wilhelm, Roithmayr, & Helmreich, 1992). Within the airline industry, Helmreich and his associates have developed a series of pre-training questionnaires—such as the Flight Management Attitudes Questionnaire (FMAQ)—to gauge pilot and flight deck attitudes towards the communications, organization, and management principles embedded in the CRM philosophy (Helmreich, Merritt, & Sherman, 1993). Before attempting any measurement of TEAM coordination, the researcher must administer carefully worded questionnaires to each crewmember before he/she receives the CRM training modules. These are represented by the ovals labeled Measurement Instrument, whose arrows feed inside the bigger box. In the next section, we describe several data collection instruments that we have developed expressly for this purpose.

Moving to the right in Figure 6, we reach the large TEAM coordination module. Within our model, there are three major links to TEAM coordination. These are Crew Background and Baseline Attitudes, discussed above, as well a box labeled Training Events. Since our emphasis is on TEAM processes and their relationship to performance, we have opted to represent Training Events as a single, undifferentiated component. However, we recognize that, in practice, there are host of salient training events during CMT. These include the CRM academics the crew receives, all relevant technical and combat tactics training, as well as the CMT mission scenario and scripted events that are presented before and after the crew flies in the WST. In the concluding section of the report, we will discuss some implications of performance-based research for modifying the conduct of CMT. At that time, we will decompose the Training Events module into its constituent components.

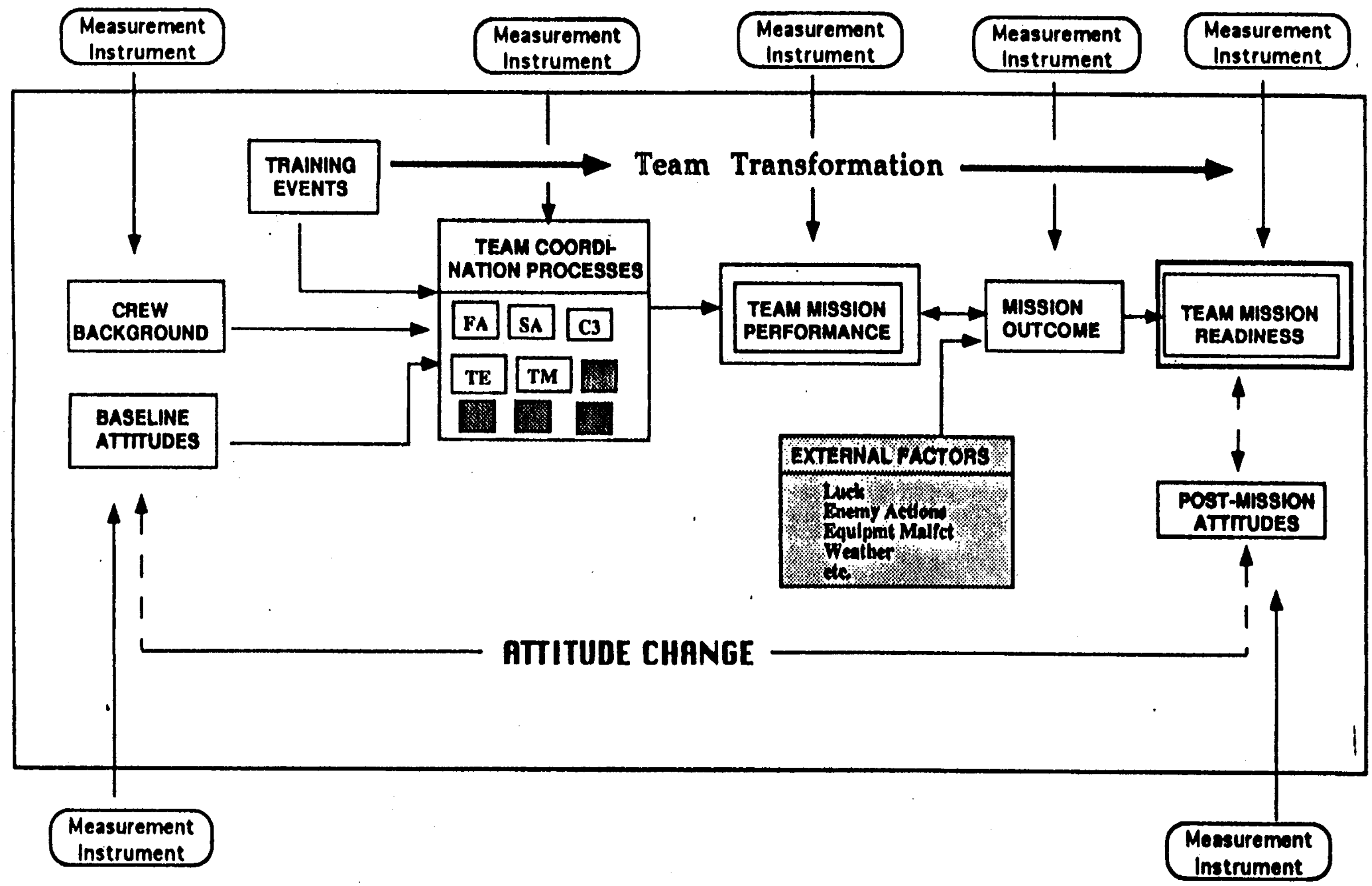

Figure 6. Conceptual Model of TEAM Performance Measurement in Combat Mission Training.

Looking inside the TEAM coordination processes module, we have included the five functional areas or subprocesses of present research interest: FA, SA, C3, TM, and TE. These subprocesses are depicted in white to indicate their coverage by the measurement instrument in the oval above. However, we reiterate that these five by no means exhaust the range of subprocesses that would fall under the larger TEAM process-coordination umbrella. Indeed, factors such as group cohesiveness, personality, group dynamics, and leadership, plus many others, are classes of variables worthy of study in their own right. We represent these as unmeasured (at least by us) factors by placing a set of gray-shaded boxes inside the larger TEAM coordination module.

Continuing on, we see that the output of the TEAM coordination module feeds directly into the TEAM Mission Performance module. However, our model makes the most sense if we first discuss the Mission Outcome module. As conceptualized here, Mission Outcome consists of those indices that, by all accounts, would be used to conclude that the TEAM's mission accomplished its stated objectives. When the mission is performed in a WST, these accounts can often be recorded by computer, and hence are free of the subjective elements that usually characterize the process side of things. For the type of scenario described in the previous section, example outcomes would include accuracy of airdrop (i.e., distance from the drop zone), percent of fuel successfully transferred, average course deviation from the planned ground track, performing infils within the prescribed control time windows, time spent on ground before the exfil, and minimizing (or avoiding) exposure to threats.

As the focus of measurement, Mission Outcome has the advantage of encompassing the criterion environment and being the ultimate yardstick in which the operations and training commands are most interested. But from a research standpoint, sole reliance on outcome is risky as there are many factors that may act to degrade outcome—ones having little if anything to do with the effectiveness the combat mission team. Some of these factors are listed in the gray-shaded box underneath Mission Outcome, such as luck, unexpected enemy actions, equipment malfunctions, weather, and the like. Unfortunately, when assessing TEAM coordination under operational or training conditions, the researcher will usually have little, or at best only partial, control over these external forces. As such, the researcher will inherit a great deal of noise and uncontrolled variability in his/her outcome measures, making it difficult (if not impossible) to infer whether good versus poor TEAM coordination had occurred.

It is for this reason that we include a TEAM Mission Performance module in our conceptual model. We place this module between the process and outcome modules since it is assumed to mediate the relationship between TEAM Coordination Processes and Mission Outcome. This mediation occurs in several forms. On the one hand, we assume that process has a fairly direct and immediate relationship to TEAM mission performance, and indeed it is this link that we most wish to capture in our research. By Mission Performance, we mean those indices that directly result from the successful (or failed) execution of important TEAM processes. In our research, TEAM Mission Performance is reflected in such indices as quality of the pre-mission briefing, completeness of the navigation chart(s), as well as SME-supplied ratings of how well the TEAM as a whole executed each phase of the mission, including low-level navigation and threat avoidance, aerial refueling, air drop, and the infil/exfil procedures.

On the other hand, mission outcome may reflect many factors, including team performance. Therefore, we connect TEAM Mission Performance and Mission Outcome with a bidirectional arrow. Some of our indices of TEAM performance will be based on data that would

normally be used to measure Mission Outcome. That is, our intent at an analytic level is to partial out the mitigating effects of the external factors as much as possible in order to arrive at a more "pure" index of TEAM performance. However, one of the prices that researchers must pay for this partialling out is greater reliance on subjective—i.e., SME-supplied ratings—data that are presently the only practical method for "filtering out" the effects of external factors. Once those unwanted factors have been screened or at least minimized, we are in a position to explore the relationship between individual elements of the TEAM coordination processes and the corresponding performance indices. As we discuss below, there have been very few attempts in the literature to accomplish such an in-depth analysis using a CRM research paradigm. However, the Povenmire et al. (1989) study stands as an impressive first step in this regard.

Moving to the far right and bottom of Figure 6, we depict the remaining modules that we wish to subject to measurement. At the far right, there is a Team Mission Readiness module that reflects the end state of the crew at the conclusion of CMT. Following the execution of the mission scenario in the WST and formal debriefing by the instructors, the aircrew should have become "transformed" into a TEAM which is ready to perform similar missions under operational conditions. While such transformations are not observed directly, they can be inferred from behavior and attitude changes noted by trained instructors as well as by tracking how well the crew performs once it returns to its operational unit (Bruce, 1989). We have placed a Measurement Instrument oval above this module to reflect our belief that one can indeed measure an aircrew's mission readiness. In the final section of the report, we discuss plans for developing such an instrument, one that can be used by squadron leadership to assess the readiness of a given aircrew before they enter the aircraft for a given mission tasking.

Below Team Mission Readiness is the ninth and final module, Post-Mission Attitudes. By comparing the crew's attitudes toward CRM (and its associated concepts) after the mission with those obtained during the pre-mission baseline, the researcher will be able to determine the degree to which their attitudes changed over the course of CMT. From a strict experimental standpoint, one cannot, of course, unequivocally attribute a change in attitude to the occurrence of CRM-oriented training. Nevertheless, a close association in time and space between CRM training and attitude change is certainly suggestive of a direct link, and is an assumption that is shared by notable experts in the area (Wilhelm, 1991; Helmreich, 1991). As we shall discuss below, a number of surveys have been conducted which demonstrate that CRM training is indeed related to positive changes in attitudes toward CRM, and the assumed acceptance and utilization of those concepts. However, it is important that one not mistake a relationship between Training Events and CRM Attitude Change for the more fundamental relationship between TEAM Coordination Processes and TEAM Mission Performance. It is this latter relationship that has the greatest implications for CMT, but unfortunately, it has been subjected to very limited empirical investigation up to this point.

Relevant Research

This subsection offers a brief summary of selected research studies which, either explicitly or implicitly, have attempted to empirically demonstrate a positive relationship between TEAM coordination processes and some measure of mission performance. In this selective review, we consider several recently reported studies by the Army, Navy, and the airlines industry, using our conceptual model to interpret the major findings and conclusions. We conclude by describing the methods and results of the Air Force sponsored studies of aircrew coordination. As will be discussed, these studies embody some of the key methodological features required to

conduct high-quality TEAM coordination research in a CMT setting, and are ones that we will attempt to emulate in our own work.

Army

Under the direction of the Army Research Institute (ARI), the Army has been conducting a series of studies to assess the effectiveness of its new Aircrew Coordination Training (ACT) program (Leedom, 1990). The impetus behind the revamped program was a response to critiques from aviators that the "existing training packages for crew coordination lacked objective standards for measuring crew performance" (p. 10). In a high level document intended to disseminate the basic concepts of the program, the Army defined crew coordination as "a set of principles, attitudes, procedures, and techniques which transforms individuals into an effective crew" (p. 10). Interestingly, the Army's focus on transformation is similar to the view espoused in the present conceptual model. In addition, and as noted previously, the Army decomposed Team Coordination processes into 13 subprocesses entitled BQs. Within the program, each BQ is defined by a set of performance standards and evaluation dimensions. Qualified IPs use the BQs to rate aircrew effectiveness during check rides in much the same manner as they assess all the other Aircrew Training Manual (ATM) technical tasks (Leedom, 1994).

For the research phase of the program, the Army tasked ARI with collecting empirical data to document the effectiveness of ACT. An initial study addressed the issue of attitude change—would aviators accept the new program? Surveys conducted at Fort Campbell showed that, indeed, aviators and IPs exhibited positive changes in attitudes toward ACT over the course of training (Zeller, 1992). While encouraging for implementation, it was acknowledged that the results do not substantiate whether the BQ-oriented program has a positive impact on either crew mission performance or mission outcome. Such a demonstration would require the conduct of further studies (Leedom, 1994).

Following introduction of the ACT program, ARI commissioned several studies to demonstrate the relationship between aircrew coordination and mission effectiveness. A study by Thornton, Kaempf, Zeller, and McAnulty (1992) is representative of the Army's methodological approach. In that study, 19 crews of two aviators each flew a combat-oriented mission in an advanced UH-60 Black Hawk simulator. The mission involved inserting troops in a cross-Forward Line of Troops (FLOT) operation into enemy territory. While several hours of planning time was given to each crew prior to the mission, no data were collected on crew coordination processes at this time. Extensive audio/video recordings were taken during the mission itself so that researchers could, after the fact, review the tapes to extract indices of crew coordination.

As viewed through our conceptual model, the Army took the following approach. First, aircrew coordination was measured exclusively in terms of communication. To that end, two researchers worked independently to develop a communications protocol that was used to assess each crew in terms of the rate, pattern, content, and quality of interactions along 13 functional categories (inquiry, command, declarative, etc.). Second, with regard to performance, mission effectiveness was defined in terms of three general functions: navigation accuracy, threat evasion, and shooting a nonprecision instrument approach. These were chosen based on their strong a priori relationship to mission success for any given mission. That is, although accurate terrain flight and complete threat avoidance do not guarantee mission success, failure on either set of tasks will ensure that the crew does not complete its mission.

Within each function, videotapes were reviewed to extract a set of reasonably objective indices. For navigation, crew performance was measured in terms of course deviations from the planned ground track and amount of time spent off-course. Threat avoidance performance was measured in terms of number of threats encountered during the mission and time exposed to each threat. Not surprisingly, these two sets of measures were highly intercorrelated as crews that navigated poorly tended, on average, to encounter more threats. For the third function, the quality of the crew's instrument approach performance was rated by two researchers using a detailed checklist derived from the supplied instrument approach plate.

Overall, there was some evidence that the crew coordination process—as defined by patterns and types of cockpit communications—showed a significant relationship to some of the mission effectiveness indices. Although rate of communication did not differentiate among the crews who performed poorly (i.e., those who navigated inaccurately, were exposed to threats, and had poor approach proficiency), there were trends in the data which suggested that certain types of communications profiles were consistently related to outcome. For example, in one analysis, the researchers reported that crews who were successful in evading threats had a pilot-flying (PF) who issued more acknowledgments than his PF counterpart in the unsuccessful crews. With regard to navigation, the researchers similarly found that crews who had higher levels of navigation accuracy acknowledged commands more frequently than did crews who navigated poorly. There was a much weaker relationship between communication patterns and instrument flight performance, although the crews that performed the approach better tended to provide more declarative information than the less successful crews.

In summarizing their results, the authors conclude that there is some evidence for a direct relationship between the communications aspect of aircrew coordination and outcome-based measures of performance. However, in interpreting their results, Thornton et al. were quite candid about the low levels of technical proficiency observed in many of the crews. Indeed, problems with map interpretation, terrain feature identification, and issuing standard radio calls were prevalent among the less successful crews. Yet, these are basic ATM tasks and are assumed to be mastered before the crew moves on to acquire the ACT/CRM skills. Hence, their results must be interpreted in this light.

In addition, it is evident that the researchers' definition and measurement of coordination was, from our standpoint, somewhat narrow since many other relevant subprocesses—SA, resource allocation, and the like—were not included. Moreover, measurements were only taken during certain parts of the mission, with the mission planning period not covered at all. Interestingly, the authors provide anecdotal evidence to support a relationship between planning quality and mission outcome. Specifically, they noted that crews that performed the instrument approach better had spent more time studying the approach plates during planning, and hence needed to refer to it less often during the high workload landing phase. Such evidence encourages our view that measurements of the TEAM coordination process should encompass the *entire* mission, including the initial briefing, mission planning, as well as each phase of the simulated mission.

Navy

A recent study by Brannick, Prince, Prince, and Salas (1995) illustrates the Navy's approach to demonstrating an empirical relationship between team coordination processes and team performance. Consistent with the present view, the Navy considers team coordination to be a set of processes. Specifically, coordination is defined as the "essence of teamwork; it is the

process, the moment-to-moment behaviors, by which interdependent team members achieve important goals" (p. 641).

Of particular interest in this study was whether the validity of some previously identified dimensions of coordination could be established by having a set of judges independently rate a large number (N = 52) of two-person aircrews flying a low-fidelity, tabletop T-44 flight trainer. The dimensions rated were the ones discussed in the previous section, and included: Assertiveness, Decision Making, Adaptability, Situation Awareness, Leadership, and Communication. Each dimension was defined in terms of behaviors that would likely occur during the mission scenario. For example, good communication involves clear and accurate sending and acknowledging of information, instructions, or commands. Based on these criteria, a set of observable behaviors was created in checklist form for each dimension. They were then used by the judges to assess crew/team coordination during the mission. This created a set of behavioral anchors for each point on a 5–point rating scale for each coordination dimension. Thus, the Navy's definition of coordination process is much broader than the Army's, and even includes some subprocesses (e.g., assertiveness, leadership) that have been omitted from our model.

Two non-tactical mission scenarios were flown by each crew. Both scenarios involved point-to-point flying, in which several unforeseen events occurred. For example, in one scenario, an "onboard passenger" had a heart attack that required an emergency landing at a divert base. To assess performance, an experienced IP rated each crew on a 20–item checklist. Items included completing flight checklist, making immediate corrections to altitude and course deviations, questioning a delay over water, and so forth. Based on a priori analysis, the IPs' assessments of these items were algorithmically transformed into an overall 5–point rating of crew performance. From an analytic standpoint, the study was designed to look at correlations between ratings of process and performance, intercorrelations among process ratings, and inter-rater reliability among independent raters of process. Interestingly, the authors had a large number of judges, 18 in all, rate the coordination process of each crew by having them watch a videotape of the mission scenario. None would be considered SMEs in the strict sense, with 12 of the raters undergraduate psychology majors.

As an added factor, the study entailed having two different crews fly the mission. One group consisted of current Navy student T-44 pilots; the other group was composed of IPs with considerably more operational experience. Experience level was included as a between-subject factor to see which coordination dimensions would exhibit significant elevations within the more experienced crews.

In support of their methodological objectives, Brannick et al. (1995) demonstrated a reasonably high degree of consistency among the raters in assigning process scores across the two scenarios. This consistency was taken as evidence for the quality of the rating instrument and the selection of concrete behaviors for direct observation. They accordingly concluded that it is possible to have judges supply internally consistent ratings by observing simulations of the mission.

The study also demonstrated a fairly strong relationship between the various team coordination process ratings and overall team performance, with the latter measure provided by the expert rater. Indeed, for both scenarios, all six process dimensions showed significant positive correlations with the expert performance score, with correlations ranging from $r = .43$ to $r = .69$. Unfortunately, the authors did not adjust their overall alpha-level (i.e., the probability of making

a Type I inference error) for the large number of statistical tests that were performed. As will be discussed in the next section, when large number of tests are performed, the preferred method of adjustment involves comparing each individual test to a more stringent (i.e., higher) level of significance (Harris, 1994).

To determine whether significant process-performance correlations would still hold under the more stringent alpha-level assumptions, we recomputed the authors' t-values (which were not reported in the article) by the formula t = r*SQRT (N-2)/SQRT ($1-r^2$). That is, since the values for r and N are known, one can solve for t for each test and compare the resulting value to standard tables of t-statistics to determine whether the reported significance levels would still hold (Hays, 1973). Given the large number of tests performed, an alpha level of .001 is more appropriate than using .05 for each test, yielding a critical t-value of 3.50 for the 49 degrees of freedom that were available for the tests (i.e., df = N-2, where N = 51, the number of crews observed). Even with this more stringent test, all but one process-performance correlation (Assertiveness/ Flexibility in the second scenario) still proved significant.

In a separate analysis, Brannick et al., (1995) compared the mean coordination ratings between the trainee crews and IP crews on each process dimension. Surprisingly, only some of the dimensions showed a significant difference in favor of the IPs. These included Assertiveness, Decision Making, and Communication. However, after applying the more stringent alpha-level to control for multiple testing, none of the mean differences proved significant. The lack of an experience effect is somewhat surprising, and it may reflect the absence of marked differences in mission performance between the two groups. Unfortunately, comparisons between IPs and trainees on the expert-supplied mission performance rating were not reported. Consequently, we may only speculate on the degree to which the non-tactical mission scenarios were sufficiently challenging to produce performance variations across crews having substantial differences in operational experience.

Airlines

For the airline industry, the vast majority of published empirical research on CRM has been performed by Helmreich, Wilhelm, and their colleagues (e.g., Helmreich et al., 1993; Law & Wilhelm, 1995; Merritt & Helmreich, 1995). Given the emphasis that commercial airlines have placed on encouraging pilots and others on the flight deck to embrace CRM concepts, it is not surprising that most studies have measured attitude change as a function of exposure to CRM seminars and line-oriented flight training (LOFT) scenarios (e.g., Helmreich & Wilhelm, 1988). In this regard, Gregorich and Wilhelm (1993) noted that whereas CRM seminars have demonstrated an "increase in targeted attitudes and motivations toward CRM concepts, there have been no links to behaviors in LOFT or to flight operations—i.e., mission performance" (p. 193). Similarly, in their survey of the aircrew coordination literature, Kaempf and Klinger (1993) concluded that the majority of CRM studies have examined whether background variables, most notably personality and crew task structure, have an impact on the coordination processes.

Compared to military R&D, the airlines shift in emphasis toward demonstrating CRM process-performance relationships has been rather slow in coming. Early on, their research looked at relationships between CRM processes, as inferred from acceptance of CRM seminars and concepts, and flight safety. Analyses of accidents and incidents were performed that, while suggestive of a relationship, were far from definitive since accident rates, as an outcome index, are sensitive to a host of external factors that may mitigate detecting subtle yet reliable trends (Diehl, 1994).

As the research base grew, it became clear that more definitive links needed to be established between engaging in CRM processes, such as clear communications and information sharing, and crew performance. Starting in 1987, NASA, University of Texas, and the FAA initiated a joint program to develop and validate an instrument that could be used by expert evaluators to rate the performance of aircrews during checkrides and LOFT simulation missions. Referred to as the Line LOS (Line Operational Simulation) Checklist (LLC), this instrument was designed as an improvement over the traditional method of pass-fail grading (Taggart, 1995). From a research standpoint, the LLC was a significant advancement as it could potentially yield data concerning the highest levels of crew effectiveness, rather than the minimum proficiency levels that are recorded during typical training evaluations (Law & Wilhelm, 1995).

However, in the initial versions of the LLC, it was discovered that airlines evaluators—not trained in psychometrics or evaluation—had difficulty using the instrument. This difficulty translated into low user acceptance, uneven reliability, lack of standardization, and reduced faith in the resulting data (Gregorich & Wilhelm, 1993). Consequently, a series of incremental improvements were made to the performance instrument, culminating in LLC version 4 (Taggart, 1995). The major modification in this latest version is the use of a "behavioral marker" approach, in which each CRM dimension to be rated on the LLC—there are 31 in all—is accompanied by concrete descriptions of exemplary behaviors. A simple 4-point rating is assigned to each of the 31 dimensions, with concrete behavioral markers provided for each.

The behavioral markers were developed based on extensive interviews, surveys, and reviews of accident and incident data. For example, one of the CRM dimensions in the Team Management and Crew Communications category is "Team concept and environment for open communications established and/or maintained." Examples of behavioral markers that appear on the LLC4 are: crewmembers listen with patience, do not interrupt or "talk over," do not rush through the briefing, making eye contact as appropriate. Under the Situation Awareness and Decision Making category, one of the CRM dimensions is "Secondary operational tasks are prioritized so as to allow sufficient resources for dealing effectively with primary flight duties." The behavioral markers for this dimension are "dealing with passenger needs, crew meals, company communications" (Law & Wilhelm, 1995).

Another enhancement in the LLC4 is the inclusion of separate ratings for each phase of flight—predeparture, takeoff and climb, cruise, and descent/approach/landing. While this requirement increases the burdens of data collection for the evaluation, it has the potential benefit of pinpointing more precisely where problems in CRM and crew coordination may reside (Law & Wilhelm, 1995). The initial data show that the LLC4 is being well-received, and is preferred by check airmen over the previous versions. There is also data to support higher inter-rater reliability, and hence greater standardization, during check ride evaluations (Taggart, 1995).

The next step in the airlines research program entails demonstrating, as in our proposed study, positive correlations between CRM processes, defined by ratings and behavioral markers, and separate indices of crew effectiveness. Because their methodological approach is still evolving, the airlines have little data to report at this point. However, Law and Wilhelm (1995) recently reported the results from two airlines' surveys that used the LLC4 to measure CRM processes as well as overall crew effectiveness. This latter index corresponds to TEAM mission performance defined within our model. Unfortunately, the methodology employed thus far has had the same evaluators provide both process and performance judgments, a procedure that may induce artificially high correlations and which may not be completely valid. Nevertheless, it is

encouraging that a substantial number of airlines is considering adopting a performance-oriented evaluation standard that will yield a rich source of data from which more specific process-performance correlations may be derived.

Air Force Studies of Aircrew Coordination

Krumm & Farina (1962). Years before the present preoccupation with CRM, Krumm and Farina conducted an investigation to determine the effectiveness of providing integrated training to B-52 pilots and navigators. Using a simulator as the training medium, the investigators compared aircrews who received the standard, separate training (control group) with an experimental group who received integrated training in a specially modified simulator. Specifically, the experimental group flew multiple full mission scenarios in which the individual crew stations were linked together.

Employing an experimental design that holds up well to this day, the investigators collected coordination process data on the pattern and rate of communication between crewmembers during selected segments of the training mission. Communication was classified into seven categories, including the source and recipient of the information, the types of acknowledgment that were given, who orders courses of action, and so forth. They also collected objective measures of performance, including navigational and bombing accuracy.

The primary hypotheses of the investigators were confirmed. First, they found that the method of training had a positive impact on coordination, as the crews who trained together had better patterns of communication than the control group. Second, and most important, Krumm and Farina (1962) noted that the quality of the communication/coordination patterns was significantly related to both navigation and bombing accuracy. For example, crews who navigated more accurately also volunteered more information. On the other hand, the rate of communication was not significantly related to the objective performance measures. This finding is consistent with the results of the Predmore (1991) study described previously.

Povenmire, Rockway, Bunecke, & Patton (1989). The Povenmire et al. (1989) study of B-52 aircrew coordination represents one of the strongest attempts in the published CRM literature to demonstrate a direct relationship between aircrew coordination processes and mission performance. As will be described below, this study embodied a number of methodological features that are particularly valuable for examining process-outcome correlations in the context of CMT. Because of these strong points, we will be using their design as a "methodological baseline" from which to launch our own research on TEAM coordination in the MC-130P weapon system. Since the Povenmire et al. study is very important to our own work, we describe its conduct and results in some detail below.

Rather than employing an explicit experimental manipulation like Krumm and Farina (1962), Povenmire and his colleagues observed seven intact aircrews as a normal part of their CMT. A high fidelity B-52 WST, with stations for each crew position, served as the platform for performing a complex, tactically realistic mission scenario. The scenario entailed conducting a long-range bombing run requiring the penetration of enemy threats, accurate dropping of bombs, and intricate navigation and maneuvers. Based on prior assessments by their squadron commander, the subject-crews were thought to span a wide range of technical and CRM proficiency.

Highly trained evaluators were used to provide assessments of aircrew coordination and mission performance, with separate sets of raters used for each measure. Specifically, one group

of evaluators assessed each crew's coordination processes using a version of the LINE/MOST worksheet by Helmreich and Wilhelm (1988). Application of this worksheet produced an overall coordination score as well as scores on the individual skill dimensions.

Independently, three other evaluators assessed each aircrew's mission performance on three main factors: bombing accuracy, threat avoidance, and technical skill. The latter factor consisted of a number of subfactors, such as maintaining appropriate altitude, performing accurate navigation, and staying within designated control times. To ensure an adequate spread of performance across crews, the researchers asked the evaluators to rank order the crews from best to worst, based on their subjective impressions of the three mission performance factors.

The study's primary analysis involved the statistical relationship between overall aircrew coordination and the crew's mission performance ranking. This relationship is depicted in Figure 7. As is evident from the figure, a strong positive relationship between the two variables was obtained, with the large correlation, r = .835, accounting for 70% of the variance, an unusually high proportion for a behavioral research study.

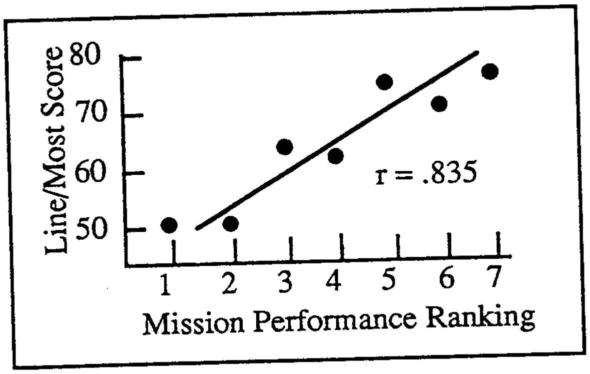

Figure 7.
Relationship Between Aircrew Coordination and Mission Performance
(from Povenmire et al., 1989).

Having convincingly demonstrated the main process-performance relationship of interest, Povenmire and his colleagues then analyzed the data further to uncover other interesting findings. In one of these analyses, they compared the experts' ratings of mission performance with the individual factors comprising mission outcome—bombing accuracy, threat avoidance, technical skill. Part-whole correlations showed that the experts mainly used bombing accuracy to make their overall judgment of mission performance, as evidenced by a significant correlation of r = .81.

In addition, the researchers performed a series of part-whole correlations on the coordination data to determine the skill dimensions in the LINE/MOST worksheet that had the highest loading with the overall coordination score. CRM dimensions that exhibited significant correlations with overall coordination included the crew's: practice of inquiry and advocacy, avoiding distractions, distributing workload, and resolving conflicts.

As part of a more detailed analysis, the researchers culled their transcribed notes carefully to extract the details of each crew's mission that tended to characterize the more effective crews. In one analysis, they examined the communications protocols and noted that the less effective

aircrews had: more conflicts, left more conflicts unresolved, and were more often in conflict about who was right rather than what was right. They also found that the more effective crews tended to confirm or challenge each other's statements more often.

Although SA was not included as a category in the LINE/MOST worksheet, Povenmire and his colleagues had their three CRM evaluators review the transcripts and independently rate each crew's SA. In each case, the correlations between the SA ranking and the LINE/MOST coordination ranking were extremely high (i.e., .86, .96, and 1.0). Consistent with the conceptual model we presented earlier, this suggests that SA is indeed a major subprocess of overall TEAM coordination.

In another analysis, Povenmire et al. reported that there was actually a negative correlation between the squadron commander's a priori assessments of crew effectiveness and the measured coordination and performance by the evaluators. However, further investigation revealed that, prior to training, there had been a substitution at the key position of radar navigator for three of the crews. As it turned out, these crews had received the highest a priori ratings by the squadron commanders. But in each case, the radar navigator had been replaced by one who was much less effective and/or aggressive.

Despite the simplicity of its design and data analysis strategy, the Povenmire et al. study stands as one of the most clear-cut demonstrations of the relationship between aircrew coordination and mission performance. Indeed, the elegance of their design is somewhat deceptive in terms of providing unusually clear insights regarding the coordination subprocesses that best predict mission performance. As will be discussed in the next section, our present research efforts will focus on expanding beyond the Povenmire work to uncover further insights regarding TEAM coordination and performance measurement. In particular, we will be describing the methodological requirements for conducting studies aimed at assessing:

- the impact of individual crewmember performance on overall TEAM effectiveness
- process-performance relationships for each mission phase separately
- effective TEAM coordination processes during the mission preparation phase
- the key behaviors that characterize the most effective TEAM subprocesses.

METHODOLOGICAL REQUIREMENTS FOR TEAM PERFORMANCE MEASUREMENT

In this section, we outline the methodological requirements to adequately measure TEAM performance within the context of CMT. While we use our own ongoing research efforts to illustrate some of these requirements, the implications of the discussion generalize beyond the specific MC-130P weapon system and its crew. We begin by addressing the major areas of inquiry that such research should address. Next, we describe five features that we believe should be incorporated into a TEAM performance experimental design. We then discuss six types of data collection instruments that will be used to obtain the measures of interest. Thereafter, we discuss some of the methodological issues that must be addressed by this type of research, and outline methods for reducing the overall technical risk. We conclude with a general strategy for analyzing the TEAM performance data that will be collected.

Areas of Inquiry

In designing our research program, a number of hypotheses concerning TEAM mission performance, TEAM coordination processes, and their relationships to each other are subjected to empirical test. Three major areas of inquiry are described below.

Does TEAM Coordination Affect Mission Performance?

The primary research topic of interest is the demonstration of a positive relationship between overall TEAM coordination and overall TEAM mission performance. While the unequivocal documentation of this linkage will provide, we believe, a valuable addition to the TEAM effectiveness literature, our conceptual approach (see Figure 6) permits us to scrutinize this relationship in further detail. For example, we may examine whether performance associated with a particular TEAM coordination area (e.g., TM) has a greater impact on overall TEAM mission performance than do the other subprocesses. Furthermore, we can assess relationships among the performance associated with TEAM coordination areas and various components of TEAM mission performance (e.g., AR or airdrop performance). In this regard, we may gauge the impact of crew SA, TM, and FA on overall mission preparation and/or performance, as well as the impact of C3 on AR performance.

To illustrate the nature of these relationships, Figure 8 depicts an idealized set of linkages between TEAM coordination and TEAM mission performance indices. In a sense, these links might correspond to hypotheses that could be subjected to empirical testing. The overall relationship between coordination and performance is shown at the top of the figure. The large, gray-shaded arrow denotes our assumption that the predictive relationship is large, but rather diffuse.

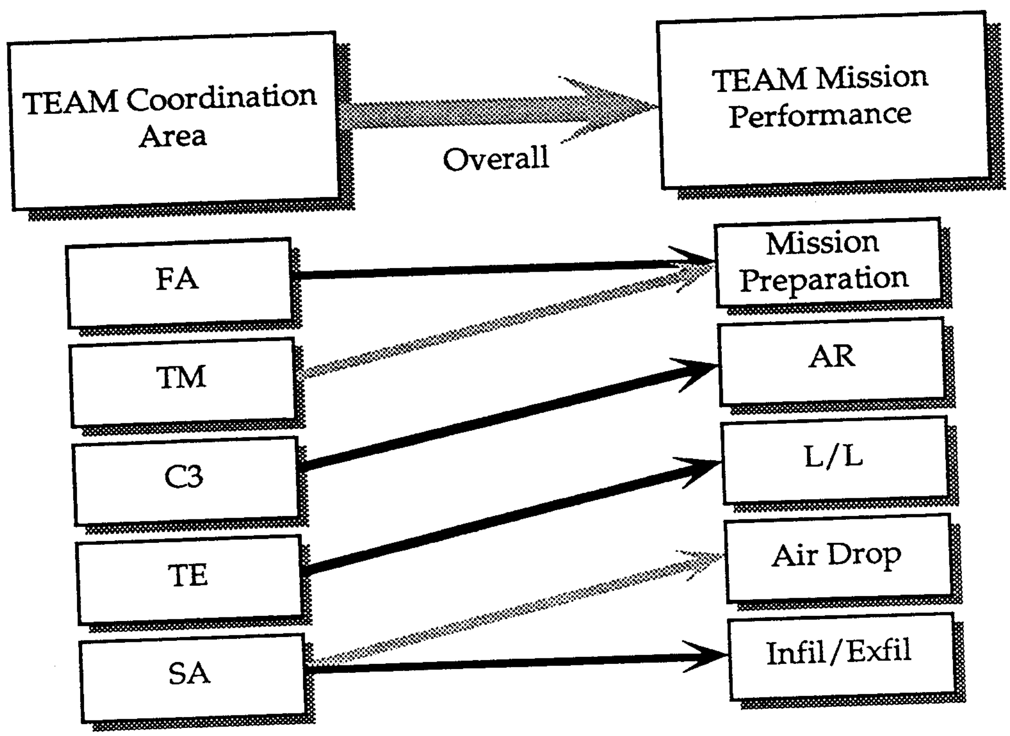

Figure 8.
"Idealized" Links between TEAM Coordination Areas and Mission Performance Indices.

Below that are displayed a set of smaller, but darker, arrows that reflect the more focused relationships we hope to find in the data. Thus, we might expect a strong, direct relationship between FA behaviors and mission preparation as well as a somewhat weaker (gray) link with TM effectiveness. Below that, we see that C3 behaviors might have a strong impact on performance during the AR phase. Similarly, TE effectiveness might exhibit a strong influence on performance in the low-level phase of the mission. Finally, SA behaviors might show a weak impact on air drop performance and a stronger influence on the mission infil/exfils. The relationships described above and depicted in Figure 8 are notional at this point, yet they are indicative of the relationships our proposed line of research hopes to explore.

Once the coordination-performance relationships have been identified, we will then want to probe the data still further to determine whether effective aircrews exhibit a consistent set of behaviors that can be "captured" and provided as feedback during training. Highly trained stan/eval personnel, instructors, and SMEs seem to be fairly accurate in distinguishing strong from weak crews; “they just know” when a crew performs well Such a response is especially true when referring to crews who work well together (i.e., have good TEAM coordination) versus crews that do not.

We will, accordingly, look beyond the statistical relationships among the process-performance data to specify more precisely and concretely those behaviors associated with the strong crews. This will be done both with regard to TEAM coordination (e.g., do strong crews clearly communicate task distribution, use checklists during planning, etc.) and TEAM mission performance (e.g., do strong crews create a drop zone sketch in addition to filling out a CARP worksheet, have above-average to exceptional mission preparation performance, etc.). Once identified, the most effective of these behavior patterns may then be fed back into MC-130P Annual Refresher Training to help mold stronger aircrews. Hopefully, some of these profiles will be sufficiently general that they may be incorporated into CMT for other weapon systems.

How Do Different Crew Positions Affect TEAM Coordination and Mission Performance?

The data matrix depicted in Figure 5 may also be expanded to isolate the behaviors associated with individual crew positions, allowing us to assess the contributions of each crew position for overall TEAM coordination and its component subprocesses, as well as overall TEAM mission performance and its scenario elements. The expansion of this matrix into a third dimension is illustrated in Figure 9. Within this overarching data structure, several questions will be addressed.

One, which crew positions have the strongest relationship to overall TEAM coordination? While it is possible that all positions are equally vital in supporting the "emergence" of an effective TEAM, the realities of operating the MC-130P may be such that some positions play a larger role than others. Thus, we may expect to see that the AC, CSO, and two navigator positions (LN, RN), by virtue of their multiple tasking and extensive communications requirements, have a larger impact on TEAM coordination ratings than do the FE or CP.

Two, an even more involved set of questions concerns the differential impact of crew position on specific TEAM coordination areas. For example, a good LN may rate high on SA and TM, but TE, FA, and C3 may not be as important. On the other hand, a good CSO may rate high on TM and C3, but lower on TE, FA, and SA. While such determinations can become somewhat involved (i.e., 6 crew positions by 5 coordination areas), they hold considerable potential for helping to identify the content of future training interventions (e.g., SA training workstation, communications checklist) that will be discussed in the last section.

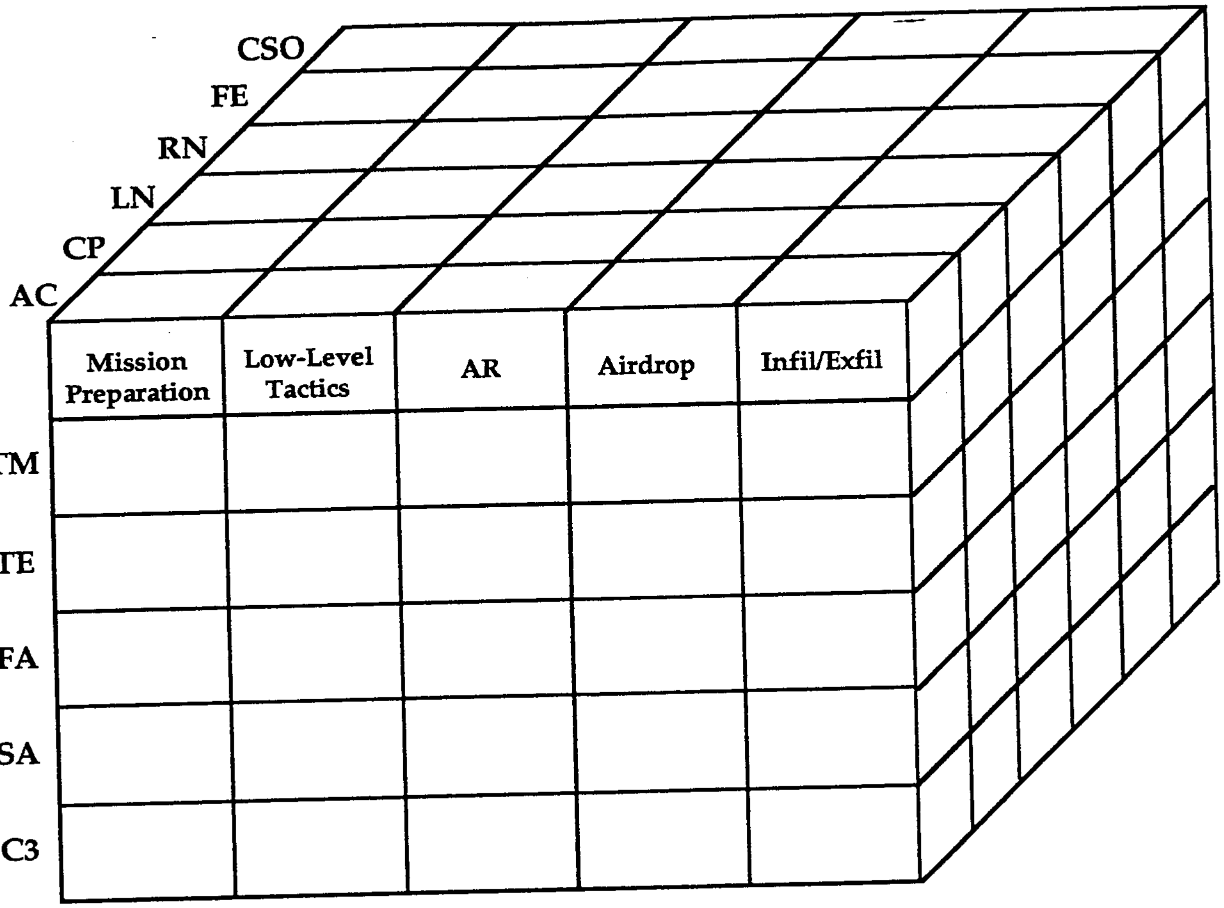

Figure 9.
Three-Dimensional Data Structure Involving
TEAM Coordination, Mission Performance, and Crew Position.

Three, do some crew positions play a larger role than others in overall TEAM mission performance? As an example, Povenmire et al. (1989) observed that the squadron-provided rankings of B-52 crew mission performance were primarily influenced by the skill level of the Radar Navigator. Based on anecdotal accounts by our project SMEs, a similar pattern may exist in MC-130P crews, with the LN's behavior being particularly central to mission performance.

Four, does the influence of specific crew positions vary when overall mission performance is divided into its specific scenario elements? For example, an AR operation is highly dependent on the ability of the FE to calculate fuel transfer and to monitor systems (e.g., hoses) during transfer. At the same time, the CSO must send, receive, and transfer messages in order to coordinate the AR with the receiving party. Moreover, one must consider the additional labor of the pilots and the navigators that is required to ensure AR success. Despite this obvious TEAM effort, it may be that due to the critical role of the CSO in coordinating this particular task, his/her performance is the most heavily weighted determinant of overall TEAM AR performance.

Are Attitudes and Experience Related to TEAM Coordination and Mission Performance?

Collecting background information on individual crewmembers permits us to explore several secondary hypotheses. These relationships are not specific experimental manipulations, but are rather post hoc groupings of crews or individuals that may further our understanding of the dynamics of TEAM mission performance and coordination. Some candidate secondary research questions are listed below:

Does squadron affiliation show different effects across phases of mission performance and coordination areas?

Do the effects of crewmember experience in the aircraft—in their current position, with the area of operation, and with current crewmembers—appear across scenario elements of mission performance and coordination areas?

Does the amount of previous CRM training and/or WST experience have an impact across scenario elements of mission performance and TEAM coordination areas?

Do the crew positions previously held influence crewmember behavior in their current position in terms of both mission performance and coordination areas?

Experimental Design Features

In this subsection, we consider the features of an experimental design that we believe necessary for a comprehensive study of TEAM effectiveness. Below, we describe five design features that we plan to incorporate into our research program.

Development of Robust TEAM Measures

Developing measures of TEAM mission performance and coordination requires that researchers focus on behaviors that are collectible, variable across crews, and operationally relevant. First, one must consider the constraints of the training environment and the resources available to determine which behaviors can be measured reliably. For example, a potentially significant navigator performance indicator may be the time navigators spend out of their seats during a mission. However, we quickly realized that, due to the lack of available space in the WST for observation, we would not be able to record this or any other non-verbal behaviors within the cockpit on a regular basis. This served as a reminder that we should focus our measurement efforts on those behaviors that we could expect, from a practical standpoint, to collect on a weekly basis. Our preliminary testing period provided insights regarding the wide variety of collectible behaviors that remained.

The second criterion is that the behaviors one selects for observation must vary across crews. For example, many training scenarios are designed so that all crews will satisfy the overall objectives of the mission—successful launch, receiving fuel during the AR, meeting the TOT, navigating within prescribed accuracy levels, shooting the designated infils and exfils, and so forth. Reliance on these mission outcome measures will pose problems for a TEAM effectiveness study as the nonvariance across crews, in essence reflecting a "ceiling effect," will yield negative results for any variable of interest. Given the overarching objective of identifying the most effective SOF aircrews, researchers must accentuate behaviors that maximally

differentiate strong from weak crews. Our preliminary testing and SME interviews provided invaluable insights regarding high payoff areas (e.g., the five coordination subprocesses) and potential behaviors on which to focus. Our continued research efforts should help to chronicle these behaviors even further.

Third, the observed behaviors must be operationally relevant. To take an extreme example, the number of times the crew rubs their eyes is a collectible and varying measure across crews. Yet its tactical relevance is highly questionable. In this regard, operational realism was one of the primary considerations in selecting the five TEAM coordination subprocesses (TM, SA, etc.) for further study. In addition, crewmembers often complain about the "soft" topics (group cohesiveness, leadership) traditionally taught in CRM courses and their weak connection to the missions crews actually fly. The TEAM coordination subprocesses we have chosen for study attempt to bring crew coordination training closer to the CMT environment, to include operationally relevant, behavioral indices of TEAM coordination. These behaviors may then be folded back into training, providing crews with immediate and relevant feedback.

Multi-Measure, Multi-Method Mix of Variables

It should be clear from the preceding discussion that the study of TEAM coordination and mission performance is a multi-faceted, multi-dimensional problem. Hence, it should not be surprising that a multi-measure, multi-method (MM-MM) mix of variables will be required to achieve a comprehensive, systematic investigation of the topic. As used here, an MM-MM approach refers to employing a battery of objective (e.g., computerized timing and counts) and subjective (e.g., ratings) measures coupled with quantitative and qualitative methods of data analysis. From an experimental perspective, a MM-MM variable mix is advisable, as it permits researchers to cast their nets broadly to tap into cognitively complex processes that may be difficult to capture with a single index. Logistically, the approach has appeal as it is fairly robust with regard to potentially devastating losses of partial data due to simulator malfunctions or subject-crew turbulence.

Referring back to Figure 6, one can see that in order to **fully** explore the links in the Conceptual Model of TEAM Performance, a minimum of seven separate measurement instrument boxes is suggested. Further, even if a researcher elects to focus on only select portions of the model (e.g., process and performance links), an MM-MM mix of variables enables one to correlate process and performance measures as well as provide opportunities to assess select intercorrelations among different types of objective and subjective performance measures.

In our research program, for example, the following methods and measures are being used: (a) TEAM coordination processes are rated and observed by an SME across the five phases of flight. This is accomplished by using headsets to monitor mission execution and from over-the-shoulder observations during mission preparation. (b) Similarly, TEAM mission performance is rated and observed by a second researcher. (c) Instructors rate individual crewmember and TEAM performance across each phase of flight. (d) Finally, select simulator performance pages are then printed out. Once obtained, we may then correlate the objective (simulator printouts) and subjective (instructor ratings) performance indices, qualitative and quantitative assessments, and process-performance relationships.

Our work is only one illustration of the amount and variety of data that may be collected when making critical TEAM mission performance and coordination comparisons. Additional

methods and measures may be fruitful, depending on the resources available in other settings. Two promising potential additions involve observations and ratings by trained observers stationed within the WST or viewing audio-visual (AV) recordings of mission preparation and execution. Observing from within, the WST is particularly advantageous because it lets researchers see a more complete set of crew behaviors, including gestures, glances, and the time particular crewmembers spend out of their seats. On the other hand, AV recordings enable researchers to replay missions repeatedly and glean more insights from the recorded data. This capability is especially important if one chooses to explore the TEAM coordination subprocess of communication (Kanki, Lozito, & Foushee, 1989; Lassiter, Vaughn, Smaltz, Morgan, & Salas, 1990).

Correlational Design—Naturalistic Observation

For our initial research efforts, we have elected to employ a naturalistic observation-correlational design rather than performing an explicit experimental manipulation. There are several reasons for this choice. First, we can take advantage of ongoing MC-130P Annual Refresher Training, using a combat mission scenario that is already in place. By working with the CMT community on a not-to-interfere basis, we have access to an experienced, inexpensive subject pool. Second, use of a naturalistic observation paradigm offers the advantages of operational relevance (external validity) and clear-cut application of TEAM mission performance and coordination principles. Third, this approach allows us the ability to immediately fold back lessons learned into the training program, without the lag time so often associated with laboratory research efforts.

Although we will not be explicitly manipulating any experimental variables in our initial work, we will nevertheless be able to examine key relationships by using post hoc groupings of the data we collect. For example, we may examine the effect of a crew's prior experience in the area of operations used in the simulated mission on TEAM mission performance. This would be done by correlating the crew's average number of hours in the area with instructor or researcher-supplied ratings of overall mission performance.

Independent Assessments of Coordination and Mission Performance

Independent assessments of TEAM coordination processes and TEAM mission performance are essential to our approach. Specifically, independent collection of coordination data from one researcher and mission performance data from a second rater helps to ensure the validity of direct comparisons and avoids the artificially inflated correlations inevitably associated with obtaining these measures from the same rater. In addition, using two or more individuals to make separate assessments—in conjunction with computer printouts—will yield more detailed accounts of both process and performance behaviors. Each rater is then able to focus all of his/her attention on the assigned dimensions.

For our research, we will be using a highly experienced, former AFSOC MC-130P operator to collect coordination data; a second researcher and four CMT instructors will make mission performance assessments. In addition, selected WST performance measurement pages will be printed out and collected at various points during the simulated mission. Consequently, process-performance relationships can be established from independent data sources, without the internal confounding from using a single rater.

Behaviorally Anchored Ratings Scales (BARS)

BARS are formal instruments that contain written descriptions of the behaviors associated with each scale value. These descriptions function as referents or anchors, and aid evaluators in determining the quality of various dimensions, such as the simple presence or absence of a behavior (e.g., prepared a mission execution checklist) or the quantitative standards that must be met (e.g., arrived within five minutes of the ARCT). Importantly, the behavioral anchors serve as criterion standards that the evaluators use in making their ratings, as opposed to making normative comparisons to other crews (or simply applying their preferences). This standardization is designed to promote reliable ratings across evaluators and crews. In recent years, the BARS methodology has become more commonly used in crew coordination research. For example, the U.S. Army has made extensive use of BARS in the evaluation of their aircrew coordination course (Grubb, Simon, Leedom, & Zeller, 1993; Grubb, Leedom, Simon, & Zeller, 1993).

Within our research program, we will apply the BARS approach in assessing TEAM mission performance and mission preparation product quality. After extensive testing and SME review and analysis, we arrived at a series of TEAM performance BARS for phases of flight, flight charts, flight plans, and the AC's mission brief. These are described in the following subsection. Perhaps in later stages of our research, when we know more about the characteristics of effective and ineffective behaviors within the various TEAM coordination subprocesses, we will expand the use of BARS to these areas as well.

Data Collection Instruments

In depicting our measurement approach, recall that Figure 6 displayed a series of "Measurement Instrument" ovals that referred to locations within the flow of training activities in which specific data collection probes could be inserted. Our intent in this subsection is to describe the six instruments we will employ for this purpose. To facilitate the discussion, we re-present our conceptual model in Figure 10, where the particular instruments are denoted in the various measurement ovals. The following paragraphs describe the purpose and scope of each instrument, along with some examples to illustrate the information items that form its content.

Crew Background Survey (CBS)

Starting at the upper left of Figure 10, the first instrument in our measurement methodology is the Crewmember Background Survey (CBS). The CBS is a self-report tool that captures relevant background information from each crewmember. The instrument's purpose is to build a descriptive profile of each crewmember's flight, mission operations, weapons system, and organizational experience. Besides requesting total flight experience information, the CBS attempts to capture each crewmember's estimate of recent flight experience with the other crewmembers attending Annual Refresher Training. This will allow us to use prior experience as a crew as a post-hoc grouping variable in some of our data analyses. This portion of the instrument is shown in Figure 11.

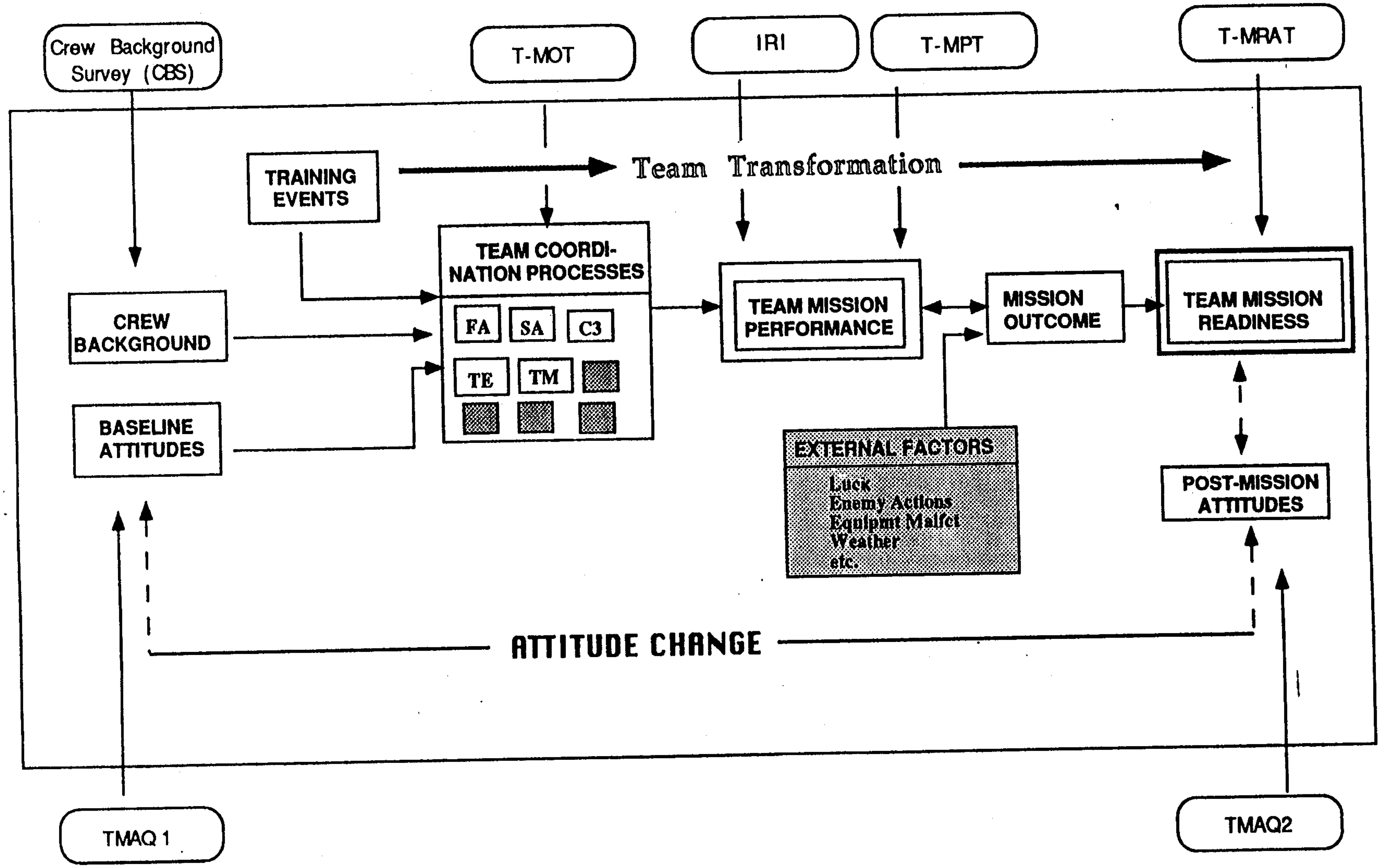

Figure 10.
Data Collection Instruments for Assessing TEAM Coordination and Mission Performance.

Circle YOUR primary crew position	NONE	LOW(Less than 75 hours)	MODERATE (75 to 99 hours)	HIGH(100 hours or more)
Pilot (AC)	None	Low	Moderate	High
Copilot (CP)	None	Low	Moderate	High
Flight Engineer (FE)	None	Low	Moderate	High
Navigator 1 (NAV1)	None	Low	Moderate	High
Navigator 2 (NAV2)	None	Low	Moderate	High
Radio (CSO)	None	Low	Moderate	High

Figure 11.
Example Portion of the CBS.

Aggregated across crews, the CBS can be used to create a similar profile at the aircrew level. This instrument is administered on the first day of Annual Refresher Training, immediately upon completing course registration, and prior to the beginning of classroom training.

TEAM Management Attitude Questionnaire (TMAQ1)—Pre-Survey

Moving to the lower left of Figure 10, the second instrument in our methodology is the TEAM Management Attitudes Questionnaire Pre-Survey (TMAQ1). This instrument is designed to collect source baseline data regarding each crewmember's attitudes toward CRM concepts and principles. The TMAQ1 is patterned after the Generation 4 Crewmember Management Attitudes Questionnaire (CMAQ) that was designed for similar purposes, but with specific application to a fully-composed flight crew of pilots and flight attendants in the airline environment (Helmreich & Wilhelm, 1988). Our approach adapted this assessment strategy, but modified the core concepts to apply to the CMT tactical environment.

In addition, the TMAQ1 was developed to satisfy a specific AFI 36-2243 (1994) assessment requirement which mandates that each MAJCOM (e.g., AFSOC) develop a method for CRM evaluation. This assessment requirement should be designed to measure how students develop, retain, and implement key coordination skills. Our proposed instrument, as a component of a larger strategy to collect comprehensive TEAM performance measurement data, satisfies the AFI evaluation requirement.

The TMAQ1 provides a systematic method of evaluating student pre-training attitudes toward CRM, and serves as a source of baseline data for assessing "crew transformation." Specifically, our conceptual model contends that some level of transformation occurs as individual mission-ready crewmembers train with other members of an aircrew team in the CMT environment. This transformation ultimately enables the mission-capable aircrew team to integrate their respective task activities and successfully complete a complex mission requiring high levels of interdependence among crewmembers. Additionally, for those aircrew teams who are unsuccessful in their CMT mission, responses to the TMAQ1 (e.g., negative attitudes towards CRM) may potentially serve as a source of explanatory information regarding the lack of information sharing among the crew).

This instrument is administered in conjunction with the CBS. An example item from the TMAQ1 is shown in Figure 12.

1	2	3	4	5
Strongly Disagree	Slightly Disagree	Neutral	Slightly Agree	Strongly Agree

1. _____ The AC should take physical control and fly the aircraft in emergency and non-standard situations.

Figure 12.
Example Item from the TMAQ1.

TEAM Mission Observation Tool (T-MOT)

The TEAM-Mission Observation Tool (T-MOT) is an integral part of our total assessment strategy. This instrument is designed to aid in recording specific individual and TEAM coordination behaviors that fall into distinct functional areas, and which are difficult to measure using questionnaires or rating forms. Measurement is accomplished using Likert scales and SME observations of critical coordination behaviors tied to a complex CMT scenario.

The T-MOT supports recording and analysis of both individual crewmember and aircrew team behaviors within the five key CRM subprocesses (TM, FA, etc.) across critical mission phases. The T-MOT structures an SME's first-hand observations of complex task performance during both mission preparation and mission execution. An internally consistent and reliable "record by exception" measurement philosophy is employed for capturing instances of extreme crew coordination behaviors demonstrated during CMT. Once collected, content analyses are performed on the recorded behaviors to permit comparisons of frequency, quality, and/or intensity across teams. This qualitative analysis supplements quantitative analyses performed on the rating data.

The primary data collection method employed in the T-MOT is over-the-shoulder structured observations, a method sanctioned by AFI 36-2243 (1994). This technique has been successfully employed in previous mission planning and mission rehearsal studies (Spiker & Nullmeyer, 1995a; 1995b). While there is a "script" to help the SME structure his observations, he is also free to record "by exception" activities, in which the observer notes attitudes, behaviors, and cognitions that seem unusually strong or weak compared to previously observed teams. This technique is principally used in observing the TEAM coordination subprocesses. When performed by a trained SME, direct observations achieve respectable levels of reliability.

The T-MOT is divided into subsections, with each subsection devoted to a particular coordination subprocess. Within each subsection, specific Yes/No checklist items are provided and assessed by the SME. The question includes an explanation field available to record notable behaviors. Example items from each subprocess are displayed in Figures 13a-13e.

(1) **Time Management** (TM): Involves the ability of the combat mission team to employ and manage limited time resources, so that all tasks receive sufficient time to be performed correctly, and critical tasks are not omitted.

1.0 An end-mission planning time should be indicated up front - most likely by an emergent "leader."

a. Did any crewmember indicate the need for an end-mission planning time? .... YES / NO
(Explain) ______

b. Was that time noted by all other crewmembers? .... YES / NO
(Explain) ______

c. Did any crewmember designate activities to establish a proper balance between their own authority, time available, and crewmember participation? .... YES / NO
(Explain) ______

d. Was adequate mission preparation time allocated for a comprehensive pre-mission briefing? .... YES / NO
(Explain) ______

Figure 13a.
Example Time Management (TM) Item from the T-MOT.

(2) **Tactics Employment** (TE): Includes all analytic activities necessary to avoid or minimize threat detection or exposure, and to successfully coordinate complex mission events and multiple mission objectives.

2.0 There are (typically) three tactical options to use in order to go undetected: Altitude, Airspeed, and Terrain.

a. Was a particular mix of tactics options considered? .... YES / NO
(Explain) ____

b. Did the crew change the tactics options as a function of difficulty in each mission phase? .... YES / NO
(Explain) ____

c. Was one option (e.g., speed) preferred over the others? .... YES / NO
(Explain) ____

d. Did any crewmember periodically review or verify the status of the threat planning strategy? .... YES / NO
(Explain) ____

Figure 13b.
Example Tactics Employment (TE) Item from the T-MOT.

(3) **Function Allocation** (FA): Includes the division of crew responsibilities so that workload is distributed among the crew, avoiding redundant tasking, task overload, and crewmember disinterest or non-involvement. Tasks should be allocated in such a manner so that crewmembers are able to share information and coordinate responsibilities.

3.0 Workload and/or task distribution should be clearly communicated and acknowledged by crewmembers.

a. Was the mission workload distribution clearly communicated and acknowledged? .... YES / NO
(Explain) ____

b. Were secondary tasks prioritized so as to allow sufficient resources for primary tasks?........ YES / NO
(Explain) ____

c. Did non-operational factors (such as social interaction) interfere with any crewmember's abilities while performing necessary tasks?........ YES / NO
(Explain)____

Figure 13c.
Example Function Allocation (FA) Item from the T-MOT.

(4) **Situation Awareness** (SA): Entails maintaining an accurate mental picture of mission events and objectives as they unfold over time and space. Emphasis and analysis are placed on the three levels of SA (perception, integration, and generation: Endsley, 1995) and their impact on team coordination.

4.0 At least one crewmember's overall SA should be high, and an assessment of mission difficulty should be made based on (for example): marginal WX, threat saturation is high, large no. of mission events, etc.

a. Did any individual crewmember indicate an overall assessment of mission difficulty? .... YES / NO
(Explain) ____

b. Did crewmember(s) prepare for unexpected or contingency situations? .... YES / NO
(Explain) ____

Figure 13d.
Example Situation Awareness (SA) Item from the T-MOT.

(5) **Command Control, and Communications** (C3): Encompasses those activities required to involve external parties in the mission, and to maintain communications with these external team members; communication within the crew; and controlling the sequence of mission events according to the mission execution plan.

5.0 Crew's willingness to challenge the system.

a. Do crewmembers request specific resources they need? .... YES/ NO
(Explain) ______

b. Do crewmembers question/challenge assumptions (e.g., within frag, threat SITREP, etc.)? .... YES / NO
(Explain) ______

c. Do crewmembers ferret out needed materials and information from all sources? .... YES / NO
(Explain) ______

Figure 13e.
Example Command-Control-Communications (C3) Item from the T-MOT.

Additionally, a 5–point rating scale (1 = lowest to 5 = highest) was developed to provide quantitative assessments across coordination subprocesses, crewmember positions, and mission phases. This technique is a fairly efficient way to generate a large amount of data within a well-defined structure in which observers use rules to assign scale ratings to some attribute of TEAM coordination. Figure 14 depicts the rating scale that was used.

Scale	Rating	Definition
1	POOR	Observed performance is significantly below expectations. This includes instances where necessary behavior was not present, and examples of inappropriate behavior that was detrimental to mission effectiveness or success.
2	MIN. REQUIRED	Observed performance meets minimum requirements but there is room for much improvement. This level of performance is less than desired for effective coordination in tactical mission operations.
3	STANDARD	The demonstrated behavior promotes and maintains coordination and mission operations effectiveness. This is the level of performance that should normally occur during tactical mission operations.
4	OUTSTANDING	Observed performance is significantly above expectations. This includes instances where necessary behavior was present, and demonstrated performance was instrumental to tactical mission success.
5	EXCEPTIONAL	Performance represents a high level of skill in the application of specific behaviors, and serves as a model for coordination, teamwork, and highly efficient tactical mission operations.

Figure 14.
A Rating Scale from the T-MOT.

Specific mission phases are also identified for assessment in the T-MOT. These phases, along with a description of their mission objectives, follow. For assessment purposes, these phases are to be considered separately during the simulator mission execution phase of Annual Refresher Training. In addition, the phases are considered in a natural sequence, as a series of mission events that occur up to, during, and immediately after the particular mission objective.

(Phase 1) **Mission preparation (MP) procedures.** The objective is to conduct premission planning and briefing activities that allow sufficient preparation of a comprehensive premission execution plan. This plan will be prepared with considerations for a medium threat environment, all major mission events and activities; and mission operations procedural constraints.

Figure 15a shows that each crewmember's and crew's demonstrated behavior during the mission preparation phase is individually rated by a trained observer, using the 1 to 5 scale, across all five coordination subprocesses.

	AC	CP	Nav 1	Nav 2	FE	CSO	Crew
1. Situation Awareness							
2. Function Allocation							
3. Tactics Employment							
4. Time Management							
5. Command, Control, & Comm.							

Figure 15a.
Matrix of Ratings Used in the Mission Preparation Segment of the T-MOT.

(Phase 2) **Low-Level (LL) tactical operations procedures.** The objective is to conduct NVG low-level flight enroute to specific mission events using proper tactical mission management procedures (altitude, airspeed, terrain masking, etc.) for a medium-threat environment.

Figure 15b illustrates one of the measurement items from the T-MOT. This one is used to record demonstration of C3 in the low-level phase of mission execution. As with the other items, the observer first indicates whether the behavior was present or absent (yes/no). He/she is then given space to provide explanatory comments.

2.0 (C3) CSO receives incoming message that (one) helicopter has ditched. The crew should spend <5 min. dealing with problem (including time for CSO to filter info.). There need not be an excessive amount of discussion about the problem's solution.

a. Was this event handled by one focal crewmember (versus a full crew emphasis)? .... YES / NO
(Explain) ________________________________
b. Did the CSO filter the message appropriately ? .... YES / NO
(Explain) ________________________________
c. Were reasonable options presented for dealing with the message? .... YES / NO
(Explain) ________________________________
d. Was an appropriate decision (outcome) ultimately concluded? .... YES / NO
(Explain) ________________________________

Figure 15b.
Example of a C3 Item from the Low-Level Segment of the T-MOT.

(Phase 3) **Aerial refueling (AR) procedures.** The objective is to successfully conduct tactical in-flight aerial refueling of (multiple) MH-53J Pave Low helicopters within prescribed time, course, and altitude constraints in a medium-threat environment.

Figure 15c illustrates another measurement item from the T-MOT. This one is used to record demonstration of TE in the AR phase of mission execution.

3.0 (TE) The AR should be completed early, so they can escape hostile airspace quicker. This also gives the crew additional flex time for later in the mission, when mission events get tight around LZ #2.

a. Did the crew exercise proper tactical refueling (phase) management procedures? .... YES/ NO
(Explain) ________________________________
b. ARCP ATA ________________ - Acceptable? .... YES / NO
(Explain) ________________________________
c. EAR Time ________________ - Acceptable? .... YES / NO
(Explain) ________________________________

Figure 15c.
Example of a TE Item from the AR Segment of the T-MOT.

(Phase 4) **Airdrop (AD) procedures.** The objective is to successfully conduct computerized automated release point (CARP) airland of special forces personnel within prescribed time, course, and altitude constraints in a medium-threat environment.

Figure 15d depicts a TE measurement item from the airdrop phase of mission execution.

4.0	(TE) Technical proficiency of airdrop should be rated by exception.	
a.	Were there any problems noted during the airdrop procedure? .... (Explain) ____	YES / NO

Figure 15d.
Example of a TE Item from the Airdrop Segment of the T-MOT.

(Phase 5) **Infil/Exfil (IE) procedures.** The objective is to successfully conduct covert infiltration and/or exfiltration at multiple tactical landing sites for transload purposes within prescribed time, course, and altitude constraints in a medium-threat environment.

Finally, Figure 15e presents another TE measurement item. This one is from the infil/exfil phase of mission execution.

5.0 (TE)	Crews can have serious problems making the approach due to poor visibility and NVG conditions.	
a.	Did the crew have problems with the approach? .... (Explain) ____	YES / NO

Figure 15e.
Example of a TE Item from the Infil/Exfil Segment of the T-MOT.

TEAM-Mission Performance Tool (T-MPT)

The first instrument in our measurement battery for collecting TEAM mission performance data is the Team-Mission Performance Tool (T-MPT). The T-MPT is designed to aid in recording demonstrated mission performance that occurs during the tactical mission preparation and execution phases. This instrument provides a structured way for trained researchers to rate the quality of individual- and TEAM-generated mission products developed during the mission preparation phase, as well as provide anchored ratings of demonstrated performance across mission phases. Within our model, the quality of the mission products developed by the aircrew team or crewmember is an index of TEAM mission performance.

Figure 16 depicts a BARS item from the T-MPT that is used to score mission flight charts developed during the mission preparation phase. The same scale is used to assess the quality of the flight charts generated by the LN, RN, and the pilots.

Figure 17 presents a BARS item from the T-MPT that is used to score AR performance. As with the planning products described above, the proficiency of AR behavior demonstrated by the aircrew team is a reliable index of TEAM mission performance.

FLIGHT CHARTS

1	2	3	4	5
- Poor. - Incomplete data. - General lack of documentation. - General quality of preparation is poor.	- Marginal. - Insufficient or inaccurate documentation. - Unaccounted for discrepancies between LN, RN, and CP charts. - Deviation plan minimally prepared. - Marginal quality.	- Adequate. - Threats plotted. - Most threat rings plotted. - Deviation plan clearly drawn. - Appropriate altitude considerations made. - Required checklist annotations made.	- Outstanding. - Threat rings plotted. - Deviation plan clearly drawn and visible for NVG conditions. - Appropriate altitude and terrain considerations made and explicitly represented in the deviation plan.	- Exceptional. - Threat contour shading provided. - Deviation plan and threat information highlighted for NVG conditions. - Documentation in excess of minimum requirements. - Threat labels.

	Score	Explain
LN Mission Chart		
RN Mission Chart		
Pilot Mission Chart		

Figure 16.
Example BARS Item from the Mission Preparation Phase of the T-MPT.

AERIAL REFUELING

1	2	3	4	5
- Poor time control to ARCP. Arrives at ARCP earlier than 1-minute prior to planned ARCT. - Poor time control to ARCP. Arrives at ARCP greater than + 1 minute (or more) late to ARCP	- Minimal time control to ARCP. - Arrives at ARCP in window of from 1-minute to 1-second earlier than planned ARCT. - Unexplained maneuvering off refueling track.	- Adequate time control to ARCP. - Arrives at ARCP in window of from 30-seconds to 60-seconds later than planned ARCT. - Maintains refueling track.	- Outstanding time control to ARCP. - Arrives at ARCP in window of from 15-seconds to 30-seconds later than planned ARCT.	- Exceptional time control to ARCP. - Arrives at ARCP in window of from on-time to 15-seconds later than planned ARCT.

Was the aerial refueling operation successful? .... YES/ NO
Explain ____

Figure 17.
Example BARS Item from the AR Phase of the T-MPT.

Instructor Rating Instrument (IRI)

The second instrument in our methodology for collecting TEAM mission performance data is the Instructor Rating Instrument (IRI). The IRI is designed to capture the unique perspectives of training instructors as they rate the demonstrated performance of their own trainees in the context of overall crew performance. For the MC-130P WST, separate instructors are used for the pilots (AC and CP), navigators (LN and RN), CSO, and FE. Each crewmember is assessed by their instructors using a Yes/No checklist and a 1-5 Likert scale. These items and ratings cover issues of demonstrated mission preparation and execution performance that are only relevant to that particular trainee's crew position. In other words, each IRI has been specifically designed and tailored to assess that crew position's roles and responsibilities throughout the mission.

Figure 18 illustrates one methodology employed in the IRI, in which pilot instructors rate their trainee using a familiar Yes/No checklist format. In addition, a comments section allows the instructor to "record by exception" those items requiring further elaboration or explanation.

PART I - MISSION PREPARATION

1. Mission Briefing	Yes	No	Comments
a. Were all appropriate issues addressed in the AC's mission brief?			

2. AC and CP Performance	Yes	No	Comments
a. Did the AC distribute the mission preparation duties appropriately across crewmembers?			

PART II: MISSION EXECUTION

1. AC and CP Performance	Yes	No	Comments
a. Did the AC manage his time well?			
l. Did the CP complete all checklists as required?			

Figure 18.
Example of AC and CP Checklist Items from the
Mission Preparation and Execution Sections of the IRI.

Secondarily, the IRI is designed to capture the instructor's unique subject matter expertise and perspective by asking for separate ratings of their individual students, as well as the aircrew team as a whole, across mission preparation and execution phases. In this manner, another index of team mission performance is provided that is independent of the assessment given by the researchers armed with the T-MPT and T-MOT.

Figure 19 illustrates a second methodology employed by the IRI. This one assigns a 1 (lowest) to 5 (highest) Likert scale value that permits the instructor to rate his/her student's performance during mission preparation and execution. The instructors are directed to rate their performance relative to other students and training they have observed from past Annual Refresher Training courses.

PART II MISSION PREPARATION

MISSION PREPARATION					
1. AC's demonstrated understanding of the mission.	1	2	3	4	5
8. The quality of the CP's Low-Level chart.	1	2	3	4	5
12. Overall mission preparation performance of the AC.	1	2	3	4	5
13. Overall mission preparation performance of the CP.	1	2	3	4	5
14. Overall mission preparation performance of the CREW.	1	2	3	4	5

Figure 19.
Example Rating Items from the IRI.

TEAM Management Attitudes Questionnaire (TMAQ2)—Post-Survey

The final instrument in our methodology is the TEAM Management Attitudes Questionnaire Post-Survey (TMAQ2). Following the line of investigation discussed under the TMAQ1, the TMAQ2 is designed to collect source data with regard to each crewmember's attitudes toward CRM concepts and principles at the **conclusion** of the week of Annual Refresher Training. As with the TMAQ1, the TMAQ2 has been patterned after the fourth generation

CMAQ (Helmreich & Wilhelm, 1988), but is used in our approach as a post-training assessment instrument.

The TMAQ2 was also developed to satisfy the specific AFI 36-2243 (1994) assessment requirement that each MAJCOM develop a CRM evaluation program. This instrument, as a component of a larger strategy to collect comprehensive TEAM performance measurement data, satisfies the AFI evaluation requirement.

The TMAQ2 provides a systematic approach to evaluating student post-training attitudes, and will serve as a source of data for assessing crew transformation. As discussed previously, our approach contends that a *transformation* occurs as individual mission-ready crewmembers train with other aircrew members in the CMT environment. This transformation ultimately enables a mission-capable aircrew team to integrate their activities and successfully complete a complex mission requiring high interdependence among crewmembers. Additionally, for those aircrew teams who are unsuccessful in their attempts, the TMAQ2 may potentially serve as a useful source of diagnostic information. By the difference between the TMAQ1 and TMAQ2 scores on individual items as well as groups of items, we will have a quantitative index of this transformation.

The TMAQ2 has been constructed using the same items as the TMAQ1, where the items have been presented in a different order to reduce the trainee's tendency to copy answers from memory. Also, the TMAQ2 has an additional section that contains questions regarding the crewmember's personal perceptions of mission readiness and mission confidence. This instrument is administered immediately upon the student's completion of his/her CMT requirements. Figure 20 depicts one of the items from this post-training survey.

I. The following questions deal with personal perceptions of crew and individual performance in the tactical mission scenario you planned for, and flew today in the Combat Mission Training course.

1 Not at all Confident	2 Somewhat Confident	3 Satisfactory	4 Confident	5 Highly Confident

1. If this mission scenario were an actual mission you had been tasked to fly, rate your confidence in the final mission execution plan that had been developed from today's mission planning phase.

Figure 20.
Example Item from the TMAQ2.

Methodological Issues

As we have previously noted, our work attempts to capture the methodological strengths of Povenmire et al. (1989) and to provide some enhancements to that seminal study. This type of TEAM performance research has many strengths as well as challenges. In this section, we describe seven major challenges for such TEAM performance research and offer our strategies for overcoming these challenges. In some cases, potentially vexing methodological issues, like having a small sample size, can actually be viewed positively if we step back and look at the "big picture."

Small Sample Size

Several factors can necessitate a small sample size when conducting TEAM performance research in realistic, operational settings. For instance, high demands placed on operational crews often preclude their participation in studies simply for the sake of research. Researchers must

accordingly develop their programs in conjunction with ongoing training efforts and acquire subjects where available. The small size of the SOF operational population is the major driver behind a small sample size. Additionally, cost, training device and instructor availability, and accommodation of reasonable timelines for research completion can all lead to the use of small sample sizes (N).

Generally speaking, there are two ways to mitigate the risk of small samples: (a) avoid N-intensive analyses, such as factor analysis or multivariate analysis of variance, or (b) look primarily for powerful effects rather than subtle ones. The level of analysis researchers choose can also help overcome the limitations associated with small sample sizes. For example, in combination with crew-based measures of process and performance, we will be collecting ratings and observations across all crew positions and several key dyads (FE and CSO) and triads (LN, RN, and AC). This will enable additional comparisons, and in some cases may increase the sample size as much as six-fold. To illustrate, when examining the relationship of a particular subprocess on individual crewmember performance, we will be able to increase the number of distinct observations from 15—the number of crews observed—to 90 (i.e., 6 crew positions x 15 subjects).

Researchers may also be able to reduce the negative impacts of a small N by making statistical adjustments based on the relative proportion of the sample to the population of interest. For example, in our research we will observe 15-20 MC-130P crews; yet this is nearly fifty percent of the total population of SOF mission qualified aircrews for this weapon system. This high proportion permits us, in turn, to calculate a reduced sample error, thereby increasing the likelihood of finding significant statistical effects, even with a small sample. Specifically, the variance of a sample mean or proportion equals its usually-calculated sample estimate, multiplied by the finite population correction, (N-n)/N-1, where **N** is the size of the population to which one wishes to generalize and **n** is the size of the sample one draws (without replacement) from that population (Winkler & Hays, 1975).

Low Crew Variability

Low variability in performance across crews is a second challenge that has plagued various TEAM coordination research studies in the past (e.g., Brannick et al., 1995). Specifically, this is the inability to find marked differences in measured performance between crews, resulting in no statistically significant effects. At the higher experience levels of the aircrew training continuum—mission qualified crews—this problem is compounded because a considerable amount of performance variation has already been removed, either through previous training or selection.

We approach this challenge from several directions. First, we plan to look at both outcome measures (mission success/failure) and mission performance measures (actual times versus estimated times, navigation accuracy, etc.). Mission performance measures are likely to exhibit more between-crew variability than simple mission outcome indices. For example, *all* crews in our study may complete the simulated mission, denying us the ability to discriminate between crews based on mission success. However, we will obtain additional measures of crew performance by phase of flight, which may yield marked differentiations between the crews. Second, we plan to assess performance using multiple measures, at multiple levels, in order to find particularly sensitive or discriminating ranges of performance. Indeed, this is one of the key design features of our research plan. We will collect quantitative assessments (ratings, counts) in conjunction with qualitative assessments (descriptions, noted exceptions), along with a rich array

of performance and process measures. This will arm our research team with sufficient data so that even subtle variations across crews may be detected.

Simulator Reliability

Whether conducting training or research in simulators, one obstacle that must eventually be confronted is simulator reliability. Advanced technology is not perfect, particularly as it involves complex computational systems. Recognizing these limitations early in designing a research study will greatly enhance one's efforts to create productive, effective programs.

We will use two strategies to minimize the deleterious effects of simulator reliability on our TEAM research program. First, we will monitor multiple phases of flight that involve elements having redundant tactical and technical significance. Within the simulation scenario, for example, there are multiple infil/exfils. If the simulator malfunctions during the course of training, we may be able to collect information on at least one of these events. Also, since our data collection strategy covers the entire mission, from preparation through debrief, we are guaranteed some data on every crew we observe, even in the extreme case of the simulator not running at all. Unlike much research on TEAM performance, we will collect TEAM process and performance data during mission preparation. Ultimately, we believe that correlations will be drawn between preparation and other phases of flight. In the event, however, that there is a complete failure of the simulator, we will still be able to make TEAM performance and process comparisons for the crew's planning session.

Second, because we are using a correlational design, we do not have the traditional idealized experimental constraint of having an equal N per condition. Pending time constraints and crew availability, we may extend our data collection period as necessary to increase our sample size and/or replace crews with incomplete data sets, i.e., those who were not able to complete the mission due to simulator malfunctions.

Instructor Rating Bias

A methodological challenge that arose during preliminary testing, as well as one noted from past research experience, is instructor reluctance to provide low performance ratings (even if deserved) for either their students or the crew. Several suggestions were offered by project SMEs as potential causes for this reluctance. The most common were: (a) pride, "this is my student"; (b) paranoia, "this is my student, and if he/she is rated poorly, it will reflect on my performance"; and (c) time pressures; providing inputs to our questionnaires is strictly voluntary, and would require additional time to justify the giving of a low rating.

We have overcome some of this reluctance by emphasizing the confidentiality and research nature of our work. We have also packaged our questionnaires in a manner that makes them quick and easy to use.

From a strategic standpoint, we have also made accommodations for instructor reluctance to provide low ratings of performance. To that end, we have designed the T-MPT for added measures of TEAM performance and will rely more heavily on the measures obtained from this instrument for performance discriminations between crews. We also have direct-access to instructors for on-the-fly interviews to supplement their ratings and plug possible gaps in the data collection "picture." From preliminary testing, we observed that instructors were more likely to comment on individual or crew performance problems rather than write them down or lower

their assessed rating on the questionnaires. We recognize instructor observations and interviews as a rich source of data, and intend to capitalize on these opportunities ~~throughout~~ throughout our research efforts.

Trainee Motivation

With regard to TEAM performance measurement, trainee motivation appears to suffer along three primary dimensions—research frequency, relevance, and task saturation. There is also a secondary dimension, simulation realism, that poses a challenge to one's research efforts.

Aircrews, and pilots in particular, are subject to an inordinate amount of research during their careers, and some may resist yet another research project conducted at their expense. The research often suffers from an aircrew's perspective because they may not realize its immediate or apparent relevance to their operational environment. Aircrews are also required to accomplish many training events and complete substantial amounts of paperwork within extremely tight schedules, and understandably, do not relish any additions to their workload.

Our research was designed with these factors in mind. We have limited the amount of direct aircrew input to our questionnaires. Our observations are designed to be as unobtrusive as possible, so as to integrate smoothly with the normal flow of training events. To that end, we work within the training environment as it exists rather than attempting to structure it into a laboratory setting. This approach provides a unique means of contributing relevant insights and feedback from our research into the CMT program of instruction.

With regard to simulation, despite the high fidelity of the MC-130P WST, aircrews often suffer from an inability to treat simulator training with the same seriousness as they treat aircraft operations. This can be compounded by instructors with similar views.

The mission scenario that is used for MC-130P Annual Refresher Training and our research program is a tactically realistic, high workload simulated mission that is representative of MC-130P SOF mission requirements. These characteristics, along with minimal instructor intervention, create a valid atmosphere for crews to prepare and execute the mission much as they would under operational conditions. In order to reduce the potential effects of instructor skepticism and variation, a researcher with high credibility among the aircrew due to his previous experience as an MC-130P navigator and instructor serves as a participant-observer. He guides the introduction of the scenario to ensure its professional and realistic presentation, and often role-plays (along with instructors) responses and directives from other agencies that would be included if it were a "real world" mission. This combination of a realistic combat scenario and the use of a participant-observer creates a serious atmosphere which closely matches the operational environment, serving to maintain trainee motivation.

Limited Access to Simulator

Due to limited available space in most WSTs, it is difficult for researchers to have direct access in making TEAM coordination and performance observations. That is, simulators are designed to match the aircraft cockpit which do not usually have additional seats for observers. The "extra" spaces that are provided in simulators usually accommodate instructors. Additionally, making over-the-shoulder observations while standing is not permitted in simulators with functioning motion systems due to safety concerns.

Several solutions to this problem may be implemented. One approach that has been used by various researchers (e.g., Thornton et al., 1992) is to set up videotape cameras within the simulator and record the mission and its participants as it unfolds. While videotaping has proved useful under some conditions, it also encounters problems. These include: filming under low light NVG conditions, cost, tasking instructors to change videotapes in the middle of the mission due to the different lengths of the mission and the videotape, and coordinating the use of two or more cameras to record larger crew sizes such as in the MC-130P.

We address the problem of limited simulator access from a pragmatic and cost-effective perspective. We have chosen to focus on the parts of the process that are reliably available, such as planning, briefings, and communication during mission execution. These are invaluable sources of data for which measurements may be consistently taken. We also will be able to print simulator performance pages through remote access to the instructor operator station (IOS). This will yield data—such as planned versus actual groundtrack, threat exposure, elapsed time, etc.—that feed directly into well-defined measures of TEAM mission performance. Nevertheless, if opportunities arise for direct observation from within the simulator (e.g., when a "jump seat" becomes available in the WST), we will of course seize upon those chances to supplement our data collection efforts.

Uncontrolled Scenario Variability

Uncontrolled scenario variability often occurs in fluid training environments, where training events, training requirements, aircraft configurations, simulator capabilities, and instructors are constantly changing. In order to circumvent this issue for our work, we have established an agreement with the 58 SOW to "freeze" the current scenario (for six months) while we collect our data. This will ensure scenario stability across the observed crews, so that valid comparisons can be drawn. The previously mentioned participant-observer, and his acceptance by the instructors, also removes some of the variability associated with scenario administration that can occur from instructor to instructor.

Data Analysis Strategy

Based on our review of the CRM research literature and the measurement model presented in the previous section, we have elected to pursue a six-step, staged strategy for analyzing TEAM performance data. This strategy will encompass both statistical and descriptive analyses, and will include checks to eliminate redundant, unreliable, or nonvarying variables. Following Harris (1994), all analyses will include appropriate Bonferroni adjustments to control for inflated experimentwise Type I error. Since the small N precludes the use of large-scale multivariate analyses (e.g., factor analysis, MANOVA), we have developed an "accordion" (collapsible/expandable) model for selecting statistical tests.

This approach is schematically represented in Figure 21. As can be seen, our approach involves looking first at the simplest (collapsed) relationships between single columns of overall process and performance data. This global analysis is intended to determine whether aircrews exhibiting more effective coordination behaviors show a concomitant higher level of mission performance. This corresponds to the two shaded columns depicted in Figure 21. This was the type of analysis that was demonstrated in Povenmire et al. (1989), and the first two steps in our data analysis strategy offer two alternative ways of making this determination.

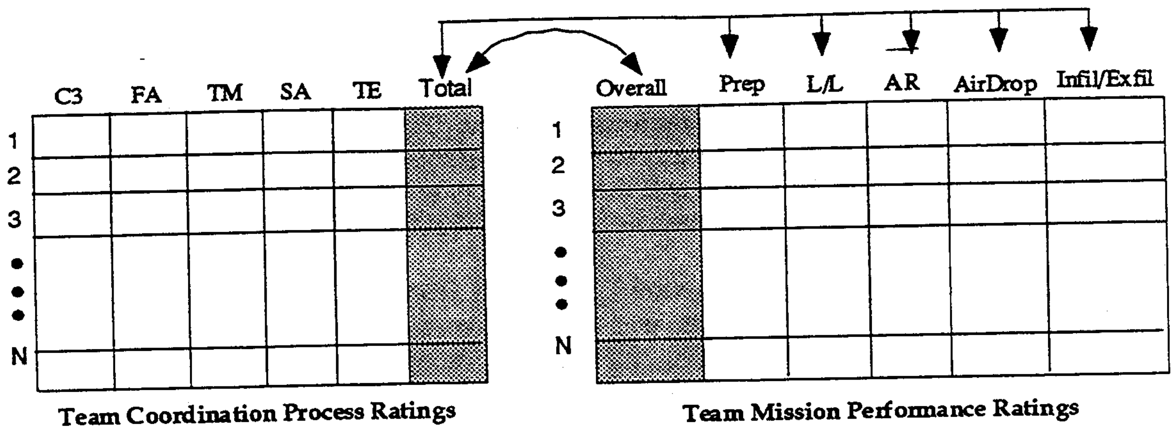

Figure 21.
Schematic Depiction of the "Accordion" Approach to Data Analysis.

If an overall relationship between process and performance is established, we then expand the data structure to consider other elements of our TEAM measurement model. As shown in Figure 21, this would involve correlating the various subprocess ratings of TEAM coordination with overall mission performance (Step 3) as well as overall TEAM coordination with the various phase-specific ratings of mission performance (Step 4). Though not shown in the figure, we may also expand the data structure across a third dimension, crew position, to correlate the relevant process and/or performance indices with one another (Step 5). Finally, we will explore some of the more detailed, one-of-a-kind relationships that may exist in the data in which combinations of qualitative variables exhibit predictive relationships with either coordination process or mission performance (Step 6).

With the above discussion as an overview, the following paragraphs describe the specifics of the analyses we plan to perform at each step. The reader should note that our descriptions of data analysis strategies are still at the conceptual level, and will undoubtedly be modified once actual data are collected. Nonetheless, the basic logic of the analysis should remain relatively stable, and importantly, the techniques described herein should apply quite well to other TEAM coordination-performance settings.

Step 1: Examine Overall Relationship Between Process and Performance

The initial analysis will entail having our two researchers independently assess their respective data packages. One researcher will have collected the TEAM coordination process data using the T-MOT. The second researcher will have collected performance data using the T-MPT. Assuming that 15 crews were assessed during CMT, each researcher will have 15 "data packages," consisting of completed instruments (either T-MOT or T-MPT), along with notes taken on Days 4 and 5 of Annual Refresher Training, and experiences gained from discussions with instructors and trainees during the sessions.

Each researcher will perform a series of pairwise comparisons in which each crew's data package is compared against every other crew's. The task itself is fairly straightforward: indicate whether Crew A or Crew B exhibited superior TEAM coordination process (or mission performance). For a sample size of 15 crews, this amounts to 105 separate pairwise comparisons (i.e., N(N-1)/2 or 15 x 14/2). Using Thurstonian scaling techniques (Guilford, 1950), one may

convert the assessed ratings into probabilities (i.e., the probability that a given crew is rated above any other crew) and then into z-scores (using the normal distribution table). Once computed, a z-score is then obtained for every crew which, importantly, has the properties of interval scale measurement rather than a simple ordinal scale. Since the two raters will have generated an interval scale measure for each crew on process and performance, one may plot each crew in two-dimensional process-performance space as shown in Figure 22. A positive relationship between process and performance is exhibited by a line trending upward and to the right.

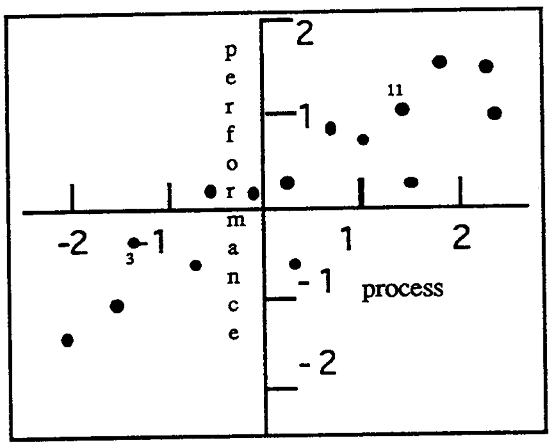

Figure 22.
Hypothetical Process-Performance Relationship Using Interval Scale Data.

Also, the interval scale properties permits one to make fairly detailed assessments of the crew's relative position in the process-performance space. In Figure 22, for example, Crew #3 has a process z-score of -1.5 and a performance z-score of -.5 (-1.5, -.5). The corresponding coordinates for Crew #11 are (1.5, 1). One may conclude that Crew #11 is 3 units higher in TEAM coordination process and 1.5 units higher on mission performance relative to Crew #3. Similar pairwise assessments can be made throughout the entire process-performance space.

Step 2: Examine Overall Process-Performance Relationship Using the Rating Data

In the second step, we will compute the Pearson product-moment correlation between the overall ratings for crew coordination taken from the T-MOT and the overall ratings for mission performance derived from the T-MPT. The size and sign of this correlation will give a statistical indication of the magnitude of this relationship. Recall that Povenmire et al. (1989) obtained a rather large correlation, .835, when comparing SME-generated rankings of crew mission performance with the LINE/MOST coordination scores. For a sample size of 7, this correlation reached significance at the .01 level (df = 5).

Assuming that our sample size is 15 crews, we will perform a simple within-subjects t-test on our obtained correlation with 13 degrees of freedom. At this point in the discussion, it is important to lay out the logic we plan to use in applying Bonferroni adjustments as a way of controlling for performing multiple significance tests. Recall from our earlier discussion that the Bonferroni technique is a fairly conservative way to keep one's overall, or experimentwise (EW), alpha level from exceeding the desired (nominal) level when a substantial number of tests will be conducted.

Typically, the adjustment is made by dividing the EW alpha by the total number of tests to be performed, thereby yielding the nominal alpha level (Harris, 1994). For example, if one plans to conduct a total of 50 statistical tests (a reasonable number for our purposes), and the desired alpha level is .05, then we have:

$$alpha_{nominal} = alpha_{EW}/50 = .05/50 = .001.$$

Thus, one must compare every statistical test against a criterion standard of .001 in order to ensure that an overall alpha level of .05 has been obtained. If we have a total of 13 degrees of freedom, then we would need to obtain a Pearson correlation of $r = .76$ to achieve statistical significance.

From our experience, an r of .76 is extremely large for any behavioral study, and is unlikely to be achieved in most circumstances. Fortunately, we have two recourses, both of which have merit.

First, if we fail to achieve overall significance, that does not mean that we cannot look further into the data for systematic relationships. Indeed, as we shall discuss below, we have further analytic options that do not depend on having achieved overall significance to be performed. This is the advantage of conducting planned comparisons rather than relying on post-hoc follow-up tests used in a factorial study, for which the overall main effect must be significant in order to proceed further. Thus, while this univariate approach to testing has the disadvantage of reduced statistical power, it has the advantage of flexibility as we are free to pursue additional significance tests based on a priori considerations derived from our conceptual model of TEAM performance measurement (Harris, 1994).

A second option, and one we will be pursuing in our work, involves distributing the statistical power (and hence the likelihood of avoiding a Type II inferencing error) **unequally** across the tests. That is, with the typical Bonferroni adjustment, all tests receive the same amount of correction (i.e., 1/number of tests) and hence the same degree of statistical power. Alternatively, one may put less of a correction on the most important tests (typically the first ones to be performed), reserving less statistical power for the remaining (and presumably less important) tests (Harris, 1994). Within our scheme, we might allocate 50% of the power for the first test, the overall correlation between process and performance. Applying this approach, then the critical Pearson r required for significance with $df = 13$ would drop to .62, a more reasonable target figure. Once this correction has been made, the rest of the statistical testing regime would proceed as planned.

Step 3: Expand Analysis to Include TEAM Coordination Areas

We will next compare the ratings from the five TEAM coordination areas with the overall TEAM performance rating. These coordination ratings will again be taken from the T-MOT, where a series of five t-tests will be performed. That is, we will compare the crew's rating on the SA, FA, TE, TM, and C3 subprocesses with their rated mission performance. The degrees of freedom will be the same as for the overall test and the assumptions similar to those described in Step 2. Accordingly, any observed r-value greater than .76 will be deemed statistically significant, with a nominal alpha of .001 and an EW alpha of .05. On an absolute basis, the analysis will indicate which areas have a statistically significant relationship to mission performance, and relatively, which coordination areas are more strongly related to performance than others.

As an aside, we will compute the correlations among all pairs of coordination area ratings to produce the 5 x 5 matrix depicted in Figure 23. As can be seen, the matrix is symmetrical (e.g., the correlation between SA and FA is the same as between FA and SA) so that only the upper half of the matrix need be represented. The variables are not correlated with themselves so the diagonal entries are also omitted, yielding 10 distinct correlations (designated r_1 through r_{10} in the figure). Inferential statistics will not be applied to these correlations, so no further adjustment to Bonferroni values are required.

	SA	FA	TM	TE	FA
SA		r_1	r_2	r_3	r_4
FA			r_5	r_6	r_7
TM				r_8	r_9
TE					r_{10}
FA					

Figure 23.
Matrix of Correlations of TEAM Coordination Area Ratings.

The primary purpose of the matrix is to provide a general perspective on the degree to which the coordination areas show high intercorrelations. To the extent that they do not, then it may be possible to conduct a Multiple Regression Analysis (MRA), in which overall mission performance is the dependent measure and the five areas are the predictor variables. Conceptually, the MRA is testing whether the slope of the line from the following equation is statistically greater than 0:

$$mp_p = B_0 + B_1SA + B_2FA + B_3TM + B_4TE + B_5FA,$$

where mp_p = predicted mission performance, B_0 = intercept, R is the multiple correlation statistic, and B_1-B_5 are the individual regression weights.

Since we may only have 14 degrees of freedom (df = N-1), an MRA is only advisable if there is a high degree of independence (i.e., non-collinearity) among the predictor variables so that the obtained regression (beta) weights are stable (Draper & Smith, 1981). However, it might be the case that the overall R statistic from the regression equation is statistically significant even if the individual r-values are not—as the former enjoys greater statistical power due to the additive effects of the individual predictor variables. Also, comparison of the individual beta weights associated with the subprocesses will give a more precise estimate of the relative impact of the coordination areas on overall mission performance.

Step 4: Expand Analysis to Include Phase-Specific Mission Performance

In Step 4, we will expand our analysis on the performance side using a similar analytic procedure as in Step 3. In this case, we will compute the Pearson product-moment correlations between each phase-specific rating of mission performance and the overall TEAM coordination rating. Thus, we will generate five distinct correlations, corresponding to mission preparation, low-level and tactics, AR, airdrop, and infil/exfil. The statistical adjustments, critical values, and degrees of freedom for these tests will be the same as those used in Step 3.

We will once again compute the intercorrelation matrix, this time among the five performance rating variables. As before, a symmetrical matrix results in which 10 distinct correlations are produced. This matrix will be reviewed to determine whether sufficient orthogonality (i.e., noncollinearity) exists to support a robust MRA in which overall TEAM coordination is used as a dependent measure and the five phase-specific performance indices are the predictors. As in Step 3, testing would be performed to determine the significance of the overall R as well as the relative magnitudes of the individual beta weights.

Step 5: Expand Analysis to Include Crew Position

Step 5 will entail the conduct of a large number of statistical tests as we expand the data structure across a third dimension to consider crew position (see Figure 9). The T-MOT is structured to permit the researcher to generate a coordination rating for each of the six aircrew positions on the MC-130P: AC, CP, LN, RN, FE, and CSO. We will first correlate each crew position's process rating with the overall performance rating to determine those crew positions having the largest impact on TEAM mission performance. The Bonferonni adjustment, degrees of freedom, and critical values of the statistical tests will be the same as those used in the previous step.

Next, we will correlate the six crew position coordination process ratings with the overall TEAM coordination rating. These correlations will indicate which crew positions have the largest impact on overall team coordination. In this regard, it will be interesting to see whether the same crew positions have the largest impact for both coordination and performance. Whereas one might suspect, at first blush, that the positions of AC, LN, and RN would receive the highest correlations in both cases, the analyses might prove otherwise. For example, it might be the case that the AC has the largest impact on TEAM coordination because of his importance in assigning functions during mission preparation. Conversely, we might find that during the mission, aircrew performance is influenced more by the proficiency of the RN and CSO due to their impact on navigation and communications.

Finally, we will examine select combinations of crew position-specific process and phase-specific ratings of mission performance. Since some 100-odd correlations could in principle be computed, selection of the correlations for testing will be guided by common sense. For example, we would be particularly interested in determining if the CSO's coordination rating correlates highly with the crew's performance rating during AR, when communications with the to-be-refueled helicopter are important. Similarly, the coordination ratings of both the LN and RN should be positively related to the crew's rated performance during the tactical low-level phase of the mission. We might also consider whether the AC's rated coordination was significantly correlated with the crew's mission performance rating. Once the analyses are actually performed, other correlations will no doubt be identified as being of interest. In each

case, though, the specific procedures underlying statistical testing will be identical to those described previously.

Step 6: Examine Other Analysis Possibilities as the Data Warrant

In the final step, we will perform a hybrid set of tests whose particulars will depend on the patterns of variation found in the data. For example, one series of analyses will involve computing partial correlations between aircrew coordination and mission performance, where squadron background has been used as a statistical covariate. Likely background variables for use as covariates would include squadron membership, amount of experience with the mission area of operations, TMAQ scores, and total amount of time flown together as a crew. Although the use of a covariate will reduce the degrees of freedom in the test, this reduction may be more than offset by the gain in statistical power brought about by reducing previously uncontrolled-for variance.

A second set of supplemental analyses will entail replicating the coordination-performance correlations performed in Step 2, in which the instructor-provided IRI data are used instead of the T-MOT. In advance of this analysis, we will examine the intercorrelations among the IRI and T-MOT data to determine the extent to which shared and unique variance exists in the correlation matrix. If the data patterns warrant, we will see whether there is sufficient variation in instructor ratings to produce a significant relationship with TEAM mission performance. If there is an overall significant effect, we may then delve into the data structure further, by considering the IRI ratings by crew position, thereby replicating portions of the analysis performed under Step 5. Since we are not sure that IRIs will be a particularly rich data source, we have relegated it to Step 6. Nevertheless, we certainly intend to explore any potential relationship with mission performance, regardless of the outcomes of the previous steps.

Finally, a third series of analyses will address the extent to which various "derived indices" of key behaviors, taken from the T-MOT checklists, can be used to predict either overall TEAM coordination or TEAM mission performance. Until actual data are collected, it is difficult to identify which indices in advance might prove fruitful. However, given our past experience in using qualitative data sets (Spiker & Nullmeyer, 1995a), we have found that careful examination of such data can prove most illuminating. To that end, we will conduct an extensive assessment to identify those indices which seem to covary most strongly with either the overall process or performance ratings. Once the leading candidates have been identified, we will then convert the qualitative indices into a quantitative surrogate so that correlations can be computed. For example, we might find that the most effective aircrews tend to exhibit some consistency in terms of: (a) having the AC serve as the leader (as opposed to some other crew position or no one), (b) making effective use of the CP during mission preparation, and (c) early recognition by one or both navigators that there is a problem in the designated control times, making the mission difficult.

Once these indices have been identified, it will be possible to construct a descriptive profile in which a crew's relative standing on each index is determined. For example, we might rank a crew either a 1, 2, or 3 on the AC index, depending on whether: the AC served as an effective leader (3), some other crewmember (e.g., LN or RN) was the leader (2), or there was no identified leader (1). Similar constructions could be developed for the other indices, yielding a crew descriptive profile or vector, such as [3, 2, 3, 1, etc.], where the dimensionality of the vector would be determined by the number of "derived indices" that were identified. This vector would then be correlated with overall TEAM coordination and TEAM mission performance.

Importantly, identification of one or more significant crew descriptive profiles might serve as the springboard for deriving a TEAM mission readiness index (see Figure 6), the prospects for which are described in the concluding section of the report.

ANTICIPATED R&D IMPACTS

We expect that the data collected from this research will provide further guidelines and recommendations which enable a greater understanding of those constituent subprocesses required for effective aircrew TEAM coordination and mission performance. The results of this investigation will produce data for CMT instructors and training developers to identify what constitutes skill in aircrew team performance protocols and procedures, thereby supporting feed-back and reinforcement procedures. Training interventions can then be developed to improve aircrew coordination subprocesses in task-specific situations where a high level of interdependency is required between members due to unpredictable or highly complex missions. The data from these instruments may also be incorporated into the CMT curriculum design cycle. Furthermore, the data may be used in the development of optimal training strategies and interventions that add incremental value to the CMT process. Finally, the produced data can be used as a baseline of comparison and application to other training program elements of a total Mission Aircrew Training System (MATS).

Ultimately, the desired goal of CMT is to produce lasting, positive behavioral change that results in improved crew effectiveness and mission performance. A successful R&D effort should validate and expand the conceptual and empirical bases of training technology, especially as applied to developing and maintaining aircrew proficiency.

We see three major impact areas resulting from the proposed research: development of an aircrew mission readiness tool, construction of improved CMT scenarios, and identification of improved CMT procedures whose effectiveness can be tested experimentally in the second year of our research program. Each impact area is discussed below.

Development of a TEAM Mission Readiness Assessment Tool (TM-RAT)

The results from the empirical study of CMT effectiveness within the confines of a WST should form the empirical foundation for the eventual development and operational validation of a TEAM Mission Readiness Assessment Tool (TM-RAT). The TM-RAT would be completed by a squadron commander or stan/eval IP to gauge a given aircrew's readiness prior to mission execution. The TM-RAT could be used for other purposes as well, including aircrew evaluation, crewmember evaluation, and instructor evaluation. The portable data collection methods used in our study will be purposely designed for incorporation into TEAM effectiveness assessments during joint exercises and other operational unit activities.

Improved Combat Mission Training Scenarios

A second area that should benefit from the research entails the eventual development of improved and more varied CMT scenarios. As we examine the study data in detail, and determine which events both challenged crews and helped them hone key tactics, we should be able to extract principles of scenario content development that could apply to the creation of more challenging, effective, and varied training events. In making these assessments, we will consider both the appropriate sequence and mix of events. We will also consider the approach

that LUTHANSA Airline is taking, which uses a Chinese menu-type strategy in which specific options are varied within a larger generic shell. This ensures variety while minimizing the amount of unnecessary design work. Such a strategy is consistent with the "vignette" approach toward scenario design that has been recommended in other contexts (Lickteig, 1991).

Improved Combat Mission Training Procedures

Finally, we fully expect that ideas for a number of procedural improvements to CMT should accrue from our research, ones that can be folded back into the Annual Refresher Training course that is serving as our research testbed. Likely areas where procedural enhancements might be found include:

- *Instructor reinforcement of key behaviors*—It is likely that we will identify ways that instructors can be more effective in providing feedback to crews, particularly in terms of providing immediate reinforcement for positive CRM behaviors. As part of this intervention, one could include specific instruction in improving an instructor's observation skills, developing an observation checklist to identify signs of poor coordination, and providing skills training in the proper use of negative feedback.

- *Consolidated checklist for planning*—Observations of multiple planning sessions should yield insights concerning the materials that optimize the TEAM's subsequent performance in the WST. We anticipate that creating a consolidated checklist of key planning events, and having the designated AC (or some other crewmember) use that list during planning to ensure completeness of the plan, will likely improve mission performance.

- *Standalone workstation for SA training*—In other contexts, we have observed the potential power in giving database developers, planners, and crews common access to the simulation environment outside the WST. In particular, such a workstation permits multiple members of the TEAM to "fly through" a geo-specific database with the aero model decoupled, so that alternate routes, tactics, and maneuvers can be quickly assessed.

- *C3 tool*—Like the planning checklist, development of a specialized C3 tool, a script-based instrument that walks the crew through a dry run of the mission, will have valuable impacts on smoothing out the "coordination wrinkles" (e.g., call signs, code words, knowing which control times are hard versus soft) that afflict most operations.

- *Selective cross training*—Specialized cross training in key crew functions (e.g., threat identification, navigation updating) might be considered as a way to improve FA during high workload periods of the mission.

- *Special tactics training*—As part of the study, we may discover some of the critical cue discriminations that aid crews in performing key tasks, such as threat evasion and avoidance. Once these cues are identified, then training in making the cue discriminations can be provided, along with potential methods for both overlearning the response as well as generalizing the response to other situations. Techniques such as above real-time training (ARTT), mediation, and guided cognitive pretraining might be considered.

- *Team training*—Determining ways to increase the amount of time the crew—and indeed the entire TEAM—spend together during training should be helpful. Presently, much of CMT is done on a duty position-specific basis, leaving crews little time to solidify the CRM concepts they learned earlier in training. Besides the crew, CMT TEAM members (or their instructor role-playing surrogates) should be incorporated, such as ABCCC, ATC, ground customers, mission commanders, logistics support, intel, other crews, and other weapon systems (e.g., helicopters).
- *Instructor acceptance of CRM*—Techniques that better engage CMT instructors in the CRM process and that help instructors accept and embrace CRM should be considered. This will be especially important during the debriefings.

REFERENCES

AFI 36-2243. (June, 1994). *Cockpit/Crew Resource Management Program.* Washington, DC: Headquarters, United States Air Force.

AFSOCR 51-30. (January, 1994). C-130 Aircrew Training. Hurlburt Field, FL: Headquarters Air Force Special Operations Command.

Andrews, D., Waag, W. & Bell, H. (1992). Training technologies applied to team training: Military examples. In R. Sweezey & E. Salas (Eds.) *Teams: Their training and performance* (pp. 283-328). Norwood, NJ: Ablex Publishing Co.

Brannick, M. T., Prince, A., Prince, C., & Salas, E. (1995). The measurement of team process. *Human Factors, 37,* 641-651.

Bruce, P. D. (1989). *Aircrew training evaluation: B-52 and KC-135 formal school training* (AFHRL-TR-88-49, AD A208 860). Williams AFB, AZ: Air Force Human Resources Laboratory Operations Training Division.

Bruce, P. D., Killion, T. H., Rockway, M. R., & Povenmire, H. K. (1991). *B-52 and KC-135 mission qualification and continuation training: A review and analysis* (AL-TR-1991-0010, AD A241 591). Williams AFB, AZ: Armstrong Laboratory, Aircrew Training Research Division.

Clothier, C. C. (1991). Behavioral interactions across various aircraft types: Results of systematic observations of line operations and simulations. In *Proceedings of the Sixth International Symposium on Aviation Psychology*. Columbus, OH.

Diehl, A. (1994, June). Crew resource management. *Flying Safety,* 8-11.

Diehl, A. E. (1992). The effectiveness of civil and military cockpit management training programs. In *Proceedings of the 45th International Air Safety Seminar*. Long Beach, CA.

Draper, N. R., & Smith, H. (1981). *Applied regression analysis.* New York: John Wiley & Sons.

Dyer, J. C. (1984). Team research and team training: State-of-the-art review. In F. A. Muckler (Ed.) *Human Factors Review* (pp. 285-323). Santa Monica, CA: Human Factors Society.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors, 37,* 65-84.

Fowlkes, J. E., Lane, N. E., Salas, E., Franz, T., & Oser, R. (1994). Improving the measurement of team performance: The TARGETs methodology. *Military Psychology, 6,* 47-61.

Gregorich, S. E., & Wilhelm, J. A. (1993). Crew resource management training assessment. In E. Wiener, B. Kanki, & R. Helmreich (Eds.), *Cockpit resource management* (pp. 173-198). San Diego, CA: Academic Press, Inc.

Grubb, G., Leedom, D., Simon, R., & Zeller, J. (1993). *Effects of crew coordination training and evaluation methods on AH-64 attack helicopter battalion crew performance* (E-1217U). Fort Rucker, AL: U.S. Army Research Institute.

Grubb, G., Simon, R., Leedom, D., & Zeller, J. (1993). *Development of candidate crew coordination evaluation methods and materials* (TR-E-21867U). Fort Rucker, AL: U.S. Army Research Institute.

Guilford, J. P. (1950). *Fundamental statistics in psychology and education*. New York: McGraw-Hill.

Harris, R. J. (1994). *ANOVA: An analysis of variance primer*. Itasca, IL: F.E. Peacock Publishers, Inc.

Hays, W. L. (1973). *Statistics for the social sciences (2nd edition)*. New York: Holt, Rinehart, & Winston.

Helmreich, R. L. (1980). Social psychology on the flight deck. In G. Cooper, M., White, & J. Lauber (Eds.) *Resource management on the flight deck: Proceedings of a NASA/Industry workshop* (NASA CP-2120). Moffett Field, CA: NASA-Ames Research Center.

Helmreich, R. L. (1991, April). *Strategies for the study of flightcrew behavior*. Sixth International Symposium on Aviation Psychology, Ohio State University.

Helmreich, R. L., Merritt, A. C., Sherman, P. K., Gregorich, S. E., & Weiner, E. L. (1993, May). *Flight management attitudes questionnaire (FMAQ)*. (NASA/UT/FAA TR 93-4). Austin, TX: University of Texas.

Helmreich, R. L. & Wilhelm, J. A. (1988). *Instructor evaluations of crew performance in LOFT/MOST* (NASA/UT TR-2-88). Austin, TX: NASA-Ames Research Center Cooperative.

Kaempf, G. L. & Klinger, D. W. (1993). *Integrated measurement of crew resource management and technical flying skills* (DOT/FAA/RD-93/26). Washington, DC: Department of Transportation/ Federal Aviation Administration.

Kanki, B. G., Lozito, S., Foushee, H. C. (1989, January). Communication indices of crew coordination. *Aviation, Space, and Environmental Medicine*, 55-60.

Krumm, R. L. & Farina, A. J. (1962). *The effectiveness of a B-52 integrated flight simulator for its coordination training potential as measured by crew communications and performance measures* (AIR-327-61-FR-239). Washington, DC: American Institute for Research.

Lassiter, D. L., Vaughn, J. S., Smaltz, V. E., Morgan, B. E. & Salas, E. (1990). A comparison of two types of training interventions on team communication performance. In *Proceedings of the Human Factors 34th Annual Meeting*

Lauber, J. K. (1987). Cockpit resource management: Background studies and rationale. In H. W. Orladay & H. C. Foushee (Eds.), *Cockpit resource management training* (NASA Conference Publication 2455). Moffett Field, CA: NASA-Ames Research Center.

Law, J. R., & Wilhelm, J. A. (1995). Ratings of CRM skill markers in domestic and international operations: A first look. In *Proceedings of the Eighth International Symposium on Aviation Psychology*. Columbus, OH.

Leedom, D. K. (1990). *Aircrew coordination training—A new approach.* Army Aircrew Coordination Training Student Guide. Fort Rucker, AL: Army Research Institute.

Leedom, D. K. (1994, May/June). *What is this thing called "crew coordination?* U.S. Army Aviation Digest, 10-13.

Leedom, D. K. & Simon, R. (1993). U.S. Army coordination training and evaluation. In *Proceedings of the Seventh International Symposium on Aviation Psychology*. Columbus, OH.

Lickteig, C. W. (1991). *Prototype methods for training and assessing future tactical command and control skills* (ARI Research Product 92-01). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Merrit, A. C. & Helmreich, R. L. (1995). Culture in the cockpit: A multi-airline study of pilot attitudes and values. In *Proceedings of the Eighth International Symposium on Aviation Psychology*, Columbus, OH.

Nullmeyer, R. T. (1995). Aircrew team training: A tutorial. From *the Seventeenth Interservice Industry Training Systems and Education Conference*. Albuquerque, NM.

Oser, R., McCallum, G. A., Salas, E., & Morgan, B. B. (1989). *Toward a definition of teamwork: An analysis of critical team behaviors* (TR-89-004). Orlando, FL: Naval Training Systems Center.

Povenmire, H. K., Rockway, M. R., Bunecke, J. L., & Patton, M. W. (1989). *Evaluation of measurement techniques for aircrew coordination and resource management skills* (UDR-TR-89-108). Williams AFB, AZ: Air Force Human Resources Laboratory, Operations Training Division.

Predmore, S. C. (1991). Microcoding of communications in accident investigation: Crew coordination in United 811 and United 232. In *Proceedings of the Sixth International Symposium on Aviation Psychology*. Columbus, OH.

Prince, C., Chidester, T. R., Cannon-Bowers, J., & Bowers, C. (1992). Aircrew coordination - achieving teamwork in the cockpit. In R. Sweezey & E. Salas (Eds.), *Teams: Their training and performance* pp. 329-353). Norwood, NJ: Ablex Publishing Co.

Prince, C., Oser, R., Salas, E., & Woodruff, W. (1993). Increasing hits and reducing misses in CRM/LOS scenarios: Guidelines for simulator scenario development. *International Journal of Aviation Psychology, 3,* 69-82.

CONTENTS (Concluded)